**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR – versi 2 - 9 Oktober 2015**

# IDENTITAS PENGUSUL

**NAMA : Ramadhan Rosihadi Perdana**

**NRP : 5112100032**

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

**DOSEN PEMBIMBING : 1. ?  
 2. Rully Soelaiman, S.Kom., M.Kom.**

# JUDUL TUGAS AKHIR

“Implementasi algoritma incremental overlapping klustering berbasis tree menggunakan teori tiga pengambilan keputusan”

# LATAR BELAKANG

Pendekatan metode klustering sering kali terbatas pada *crisp* klustering , dimana objek hanya termasuk dalam satu cluster. Sementara itu dalam praktiknya di lapangan terdapat banyak kasus dimana objek termasuk dalam lebih dari satu cluster. Selain itu , pendekatan klustering yang sudah ada biasanya menganalisa dataset statis dimana objek tetap tidak berubah setelah diproses. Padahal banyak dataset yang dimodifikasi secara dinamik, yang berarti pola yang sebelumnya dipelajari harus diperbarui.

Pada tugas akhir ini, penulis mencoba mengimplementasikan metode *incremental overlapping* klustering untuk mengatasi permasalahan tersebut.Tree dibangun dari titik-titik representatif yang dapat meningkatkan relevansi hasil pencarian.Model kluster tumpang tindih direpresentasikan oleh teori tiga opsi keputusan dan strategi tiga opsi keputusan ini didesain untuk memperbarui cluster ketika data bertambah. Metode ini dapat menentukan jumlah cluster selama proses. Hasil percobaan menunjukan bahwa metode ini dapat mengidentifikasi cluster dalam bentuk arbitrary dan tidak mengorbankan waktu komputasi dan hasil dari perbandingan menunjukan bahwa kinerja metode ini lebih baik jika dibandingkan dengan kebanyakan algoritma yang lain pada sebagian besar kasus

# RUMUSAN MASALAH

1. Bagaimana mengimplementasikan metode incremental overlapping klustering berbasis tree menggunakan konsep pengambilan tiga opsi keputusan?
2. Bagaimana hasil kinerja metode incremental overlapping klustering berbasis tree menggunakan konsep pengambilan tiga opsi keputusan?

# BATASAN MASALAH

1. Implementasi algoritma menggunakan Bahasa pemrograman C++
2. Dataset yang diujikan adalah dataset yang bersifat incremental dan overlapping

# TUJUAN PEMBUATAN TUGAS AKHIR

1. Mengimplementasikan metode incremental overlapping klustering berbasis tree menggunakan konsep tiga opsi keputusan.
2. Menganalisis hasil kinerja metode incremental overlapping klustering berbasis tree menggunakan konsep tiga opsi keputusan.

# MANFAAT TUGAS AKHIR

Tugas Akhir ini diharapkan dapat memberikan solusi permasalahan overlapping klustering untuk dataset yang bertambah jumlahnya tiap satuan waktu.

# TINJAUAN PUSTAKA

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

1. Incremental Overlapping Klustering

Incremental Overlapping Klustering adalah proses menentukan cluster pada suatu data yang jumlahnya bertambah tiap satuan waktu dan objek yang di cluster dapat termasuk pada lebih dari satu cluster. Terdapat beberapa penelitian yang merumuskan metode untuk incremental klustering atau overlapping klustering, namun sedikit penelitian yang mencoba merumuskan gabungan dari kedua model klustering tersebut.

Pada tahun 1998, untuk pertama kalinya dirumuskan metode yang menangani permasalahan incremental klustering bernama Incremental DBSCAN dengan berpacu pada algoritma klustering DBSCAN [1] . Kemudian pada tahun 2009 terdapat penelitian yang mengklaim memiliki performa 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 klustering baru berdasarkan jarak pada bentuk kluster yang tidak beraturan, namun algoritma pada tahapannya perlu untuk menscan keseluruhan database [3]. Satu tahun kemudian,pada tahun 2013 dirumuskan algoritma incremental klustering baru yang dapat memaksimalkan hubungan jarak antar pola pada kluster yang sama [4]. Dan pada tahun 2014 dipublikasikan algoritma incremental co-klustering yang menghitung pembagian objek dan pembagian fitur secara simultan, namun tidak dapat menemukan overlapping cluster [5].

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

1. Three-way decision klustering  
    Three-way decision klustering adalah metode klustering yang mengambil keputusan untuk pembentukan kluster dengan tiga opsi, yaitu :
   1. Menerima sebuah objek sebagai bagian dalam sebuah kriteria jika derajat kecocokan kriteria objek tersebut lebih besar atau sama dengan level yang ditetapkan.
   2. Menolak sebuah objek sebagai bagian dalam sebuah kriteria jika derajat 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 klustering digunakan untuk overlapping klustering sedangkan untuk crisp klustering metode yang dipakai dikategorikan sebagai two-way decision klustering. Berbeda dengan crisp klustering yang memodelkan klusternya sebagai himpunan dari objek-objek, pada model three-way decision klustering ini kluster di representasikan dalam bentuk interval.

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

# RINGKASAN ISI TUGAS AKHIR

Dalam tugas akhir ini akan dilakukan implementasi algoritma incremental overlapping klustering berbasis tree dengan menggunakan metode tiga opsi keputusan. Ketika terdapat dataset yang bertambah jumlahnya tiap satuan waktu, proses pembentukan kluster menjadi lama apabila yang diproses adalah keseluruhan data. Oleh karena itu dibutuhkan algoritma yang menyelesaikan permasalahan tersebut, dimana proses pembentukan kluster ketika data baru ditambahkan didasarkan pada informasi kluster yang telah dibentuk dan data baru yang masuk sehingga menghemat waktu karena tidak perlu mengolah keseluruhan dataset lagi.

Tahapan awal untuk menyelesaikan permasalahan ini adalah dengan mengkluster data awal, dilakukan dengan menggunakan tiga opsi pengambilan keputusan berdasarkan titik-titik yang merepresentasikan area dimana nilai kepadatannya lebih besar sama dengan nilai yang nanti ditentukan. Kemudian membuat pohon pencarian dari data yang telah diambil titik-titik representatifnya, dan ketika ada data baru masuk kita ambil titik-titik representatifnya untuk dicari relasinya dengan titik-representatif terdahulu yang kemudian digunakan untuk memperbaharui operasi-operasi pada struktur pohon pencarian.

Selanjutnya adalah mengkluster data yang baru ditambahkan. Setelah menemukan titik-titik representatif pada data baru, dilakukan pencarian dan pembaharuan terhadap graph yang dibuat pada saat mengkluster data awal. Kemudian akan dilakukan pengukuran kualitas pada kluster yang telah dibuat dengan menghitung akurasi, *F-measure* [6]*,* dan *NMI (Normalized Mutual Information)* [7]*.*

# METODOLOGI

## Penyusunan proposal tugas akhir

Proposal tugas akhir ini berisi tentang deskripsi pendahuluan dari tugas akhir yang akan dibuat. Pendahuluan ini 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 dipelajari sejumlah referensi yang diperlukan dalam pembuatan aplikasi.

## Implementasi Algoritma

Dalam tugas akhir ini nantinya akan menghasilkan program dalam Bahasa C++ untuk mengkluster dataset yang bertambah jumlahnya tiap satuan waktu. Masukan data berupa dataset yang diperbarui tiap interval yang telah ditentukan. Berikut beberapa hal yang diperlukan dalam implementasi :

1. IDE menggunakan Dev C++

## Pengujian dan evaluasi

Tahap pengujian dan evaluasi akan dilakukan dengan mengimplementasikan metode pada tugas akhir ini pada dataset dinamik dan bersifat overlapping.

## 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 | 2016 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Januari | | | | Februari | | | | Maret | | | | April | | | | | Mei | | | | Juni | | | | Juli | | | |
| 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 klustering 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 klustering 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 klustering 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-klustering: off-line and incremental approaches,” *Data Mining Knowl,* vol. 1, pp. 31-64, 2014. |
| [6] | C. A. B. Larsen, “Fast and effective text mining using linear-time document klustering,” *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM,* p. 16–22, 1999. |
| [7] | J. G. A. Strehl, “Cluster ensembles—a knowledge reuse framework for combining multiple partitions,” *J. Machine Learn. Res. 3 ,* p. 583– 617, 2003. |
| [8] | W. Wang, “A Two-Layer Night-Time Vehicle Detector,” 2015. |
| [9] | 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]. |
| [10] | M. Lie, “Gofat Wordpress,” [Online]. Available: https://gofat.wordpress.com/2012/04/12/face-detection-dengan-metoda-haar-cascade/. [Accessed 2015]. |
| [11] | 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]. |
| [12] | “Thesis Binus,” 2015. [Online]. Available: http://thesis.binus.ac.id/Asli/Bab2/2007-3-00373-MTIF%20Bab%202.pdf. |