 **DEPARTEMEN INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI DAN KOMUNIKASI**

**INSTITUT TEKNLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# IDENTITAS PENGUSUL

**NAMA : Setyassida Novian Putra Damara**

**NRP : 05111440000024**

**DOSEN WALI : Adhatus Solichah Ahmadiyah S.Kom, M.Sc**

**DOSEN PEMBIMBING : 1. Bagus Jati Santoso, S.Kom., Ph.D.**

**2. Royyana Muslim Ijtihadie S.Kom, M.Kom., Ph.D**

# JUDUL TUGAS AKHIR

“Pengolahan *Top-k Dominating Queries* Pada Data *Streaming* Terdistribusi”

# LATAR BELAKANG

Selama beberapa tahun terakhir ini, *preference queries*, kueri yang hanya mengambil *preferable* objek data dari *multidimensional* dataset mendapat perhatian yang signifikan dari para peneliti basis data. Jenis kueri ini memberikan berbagai macam keputusan multi kriteria membuat aplikasi dengan banyak manfaat, misalnya, multimedia retrieval, web search, analisis pasar, dan e-commerce. Dua metode yang paling banyak digunakan dalam *preference queries* adalah top-k dan skyline.

Diberikan fungsi ranking (penilaian) : *d* → , di mana *d* adalah jumlah atribut (dimensi), kueri dengan metode top-k mengambil objek data k dengan nilai terbaik. Keuntungan dari query top-k adalah pengguna dapat mengontrol jumlah objek data pada hasil ditetapkan dengan menentukan parameter *k*. Kelemahan dari metode ini adalah tidak selalu mudah bagi pengguna untuk menentukan peringkat yang sesuai fungsi.

Metode skyline mengatasi kelemahan ini karena kueri ini tidak memerlukan fungsi peringkat apapun. Hasil query skyline terdiri dari objek data yang tidak didominasi oleh objek data lain dalam dataset tertentu. Hubungan dominasi didefinisikan sebagai: diberi dua data *oi* dan *oj*, *oi* mendominasi *oj* jika *oi* tidak lebih buruk daripada semua atribut *oj* dan lebih baik dari pada *oj* setidaknya satu atribut. Dengan metode skyline query, pengguna Bisa mendapatkan objek data yang tidak lebih buruk dari data yang lain. Namun, pengguna tidak bisa mengendalikannya ukuran dari hasil, yang mungkin dapat mengembalikan hasil data yang sangat banyak.

Kueri top-k dominating adalah metode kueri yang menggabungkan kelebihan dari metode top-k dan skyline. Lebih spesifik, kueri top-k dominating tidak memerlukan fungsi peringkat yang didefinisikan oleh pengguna dan dapat mengontrol ukuran himpunan hasil. Atas-k Kueri yang mendominasi mengambil objek data *k* yang mendominasi jumlah tertinggi objek data dalam dataset tertentu. Artinya, nilai data objek o adalah jumlah objek data yang o mendominasi. Kueri teratas mendominasi identifikasi yang paling penting objek data secara intuitif. Hal ini dapat membantu banyak aplikasi termasuk contoh di atas.

# RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam tugas akhir ini dapat dipaparkan sebagai berikut:

1. Bagaimana cara mengolah *top-k dominating query* pada data *streaming* terdistribusi?
2. Bagaimana cara mengurangi biaya komputasi dan penyimpanan pengolahan t*op-k dominating query* pada data *streaming* terdistribusi?

# BATASAN MASALAH

Permasalahan yang dibahas dalam tugas akhir ini memiliki batasan antara lain:

1. *Dataset* yang digunakan adalah
2. Teknologi yang digunakan adalah Matlab.

# TUJUAN PEMBUATAN TUGAS AKHIR

Tujuan pembuatan tugas akhir ini antara lain:

1. Melakukan analisis dan mendesain algoritma berserta struktur data untuk mengolah *top-k dominating queries* pada data berbasis kelompok.

# MANFAAT TUGAS AKHIR

Manfaat dari pembuatan tugas akhir ini adalah:

1. Meneliti permasalahan dalam pengolahan top-k dominating query pada data streaming terdistribusi.
2. Mengusulkan dua algoritma yang efisien untuk pengolahan top-k dominating query pada data streaming terdistribusi, dengan harapan kedua algoritma ini mengurangi biaya komputasi secara signifikan.
3. Mengusulkan algoritma perkiraan dalam pengolahan top-k dominating data computation, dengan harapan dengan algoritma perkiraan ini dapat mengurangi biaya komputasi jika dibandingkan dengan algoritma exact.

# TINJAUAN PUSTAKA

## Top-K Queries

## Skyline Queries

## Top-K Dominating Queries

# RINGKASAN ISI TUGAS AKHIR

Tugas Akhir ini disusun untuk menangani masalah besarnya biaya komputasi dan biaya penyimpanan dalam pengolahan *top-k dominating query* pada *data* *streaming* terdistribusi. Pada data ini saya penulis mengadopsi metode *time-based sliding window*[1], dimana dihasilkan objek data dalam W waktu dari waktu sekarang yang akan menjadi data pantauan. Dalam proposal Tugas Akhir ini penulis mengusulkan 2 metode pendekatan untuk exact top-k dominating data monitoring. Pertama adalah pendekatan berdasarkan penyaringan dimana dalam metode ini menggunakan data yang meiliki nilai dominasi tinggi sebagai penyaring untuk menhindari pengiriman data yang tidak perlu. Kedua dengan metode cache-based dengan harapan dapat mengurangi biaya komputasi secara signifikan. Dengan memanfaatkan kedua metode ini penulis mengusulkan sebuah algoritma untuk pengolahan top-k dominating query pada data streaming terdistribusi. Untuk mengurangi biaya komunikasi dan komputasi, sebisa mungkin kita harus menghindari perhitungan yang berulang dari data top-k dominating. Pada penelitian ini, penulis mengusulkan metode lower- and upper-bounding untuk memberikan nilai untuk tiap objek data yang ada, dengan menggunakian metode ini penulis akan secara cermat untuk memilih kandidat data yang diberikan dari total keseluruhan data. Penulis juga mengusulkan metode sampling-based approximate untuk perhitungan data top-k dominating dengan harapan dapat mengurangi biaya komunikasi dan komputasi secara signifikan dengan menjaga tingkat akurasi yang tinggi.

# METODOLOGI

## Penyusunan proposal tugas akhir

Proposal tugas akhir ini berisi tentang penjelasan mengenai pendahuluan dari tugas akhir yang akan dibuat. Pendahuluan ini terdiri dari hal yang melatarbelakangi tugas akhir, rumusan masalah yang diangkat, batasan masalah yang ada, tujuan dan manfaat dari tugas akhir ini. Selain itu, dijabarkan pula tinjauan pustaka yang digunakan sebagai referensi pendukung dalam pembuatan tugas akhir.

## Studi literatur

Pada studi literatur ini, akan dipelajari beberapa referensi yang akan diperlukan untuk membantu mendesain algoritma untuk mengolah *top-k dominating query* pada data berbasis kelompok*.* Secara garis besar, ada tiga metode/teknik/algoritma yang akan menjadi pilar dalam tugas akhir kali ini. Yaitu, Top-K queries, Skyline queries, dan yang terakhir adalah gebungan dari kedua metode sebelumnnya dengan nama *Top-k Dominating Queries.*

## Analisis dan desain perangkat lunak

Pada tahap ini, penulis akan menganalisa masalah yang ada dalam pengolahan *top-k dominating query* pada data *streaming terdistribusi* dan mendesain algoritma yang dapat mengatasi permasalahan tersebut.

## Implementasi perangkat lunak

Perangkat keras yang digunakan adalah perangkat keras yang mendukung graphical processing unit (GPU) NVIDIA GeForce. Sementara, sistem operasi yang digunakan adalah berbasis Windows demi kemudahan instalasi berbagai macam perangkat lunak yang dibutuhkan. Pada tugas akhir kali ini direncanakan penggunaan perangkat keras dan sistem operasi mempunyai spesifikasi sebagai berikut:

* + - Intel Core i3-2330M 2.2GHz
    - NVIDIA GeForce GT 920m 1GB
    - 4GB of RAM
    - Windows 10 64-bit

jika kemudian ternyata digunakan perangkat keras yang mempunyai spesifikasi yang berbeda, rincian spesifikasinya akan dilaporkan kemudian pada saat penyusunan laporan akhir pada tugas akhir kali ini.

Sedangkan untuk *environment* perangkat lunak, akan digunakan bahasa pemrograman Matlab dengan berbagai macam library pendukung.

## Pengujian dan evaluasi

Pengujian dilakukan untuk mengetahui apakah system yang dibangun dengan metode yang diusulkan telah bekerja dengan baik dan efisien atau belum. Pengujian dalam algoritma ini akan dilakukan dalam beberapa cara, antara lain:

1. Pengujian Keberhasilan Algoritma

Pengujian ini berfokus pada keberhasilan dalam algoritma pengolahan *top-k dominating query* pada data streaming terdistribusi.

1. Pengujian …

## 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** | **2017** | | | | **2018** | | | | | | | | | | | | | | | | | | | | |
| **Desember** | | | | **Januari** | | | | **Februari** | | | | **Maret** | | | | **April** | | | | **Mei** | | | | |
| Penyusunan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| Perancangan Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| Pengujian dan Evaluasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| Penyusunan Buku |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | M. Kontaki, A. N. Papadopoulos and Y. Manolopoulos, "Continuous Top-k Dominating Queries," *IEEE Transactions on Knowledge and Data Engineering,* vol. 24, no. 5, pp. 840 - 853, 2012. |
| [2] | N. Zhang, C. Li, N. Hassan, S. Rajasekaran and G. Das, "On Skyline Groups," *IEEE Transactions on Knowledge and Data Engineering,* vol. 26, no. 4, pp. 942 - 956, 2013. |