

# BAB 1 PENDAHULUAN

## 1.1. Latar Belakang

Peningkatan volume, variasi, dan kecepatan data yang dikenal sebagai "Big Data" mendorong perubahan yang signifikan dalam berbagai sektor, termasuk bisnis, kesehatan, pendidikan, dan industri. Hal tersebut mendorong berkembangnya beragam jenis data yang mengubah cara kita memahami dan menginterpretasikan data. Jika dianalisis dengan benar, data ini dapat memberikan wawasan yang berharga yang memungkinkan kita untuk mengetahui lebih banyak tentang fenomena yang terjadi dan mendapatkan pengetahuan baru[19]. Dengan semakin berkembangnya kebutuhan akan analisis data dalam berbagai bidang dan bertambahnya jenis data, salah satunya data *streaming* yang dapat berguna dalam memenuhi kebutuhan pengambilan keputusan secara *real-time*[2]. Data *streaming* merupakan data kontinu *time series* yang terus menerus dihasilkan secara *real-time*[17]. Memanfaatkan data *streaming* untuk proses analisis memungkinkan pengambilan keputusan yang cepat dan akurat, sehingga pengambilan keputusan lebih responsif dan tepat[22].

Analisis data *streaming* digunakan tidak hanya di bidang industri dan bisnis, namun juga dalam memonitor dan menganalisis data komputer paralel atau yang sering disebut sebagai *High Performance Computing*(HPC). Analisis ini memungkinkan pemantauan terhadap sistem HPC secara *real-time*, sehingga kejadian seperti malafungsi dan peningkatan sistem dapat ditangani dengan cepat dan responsif, guna pengoptimalan sistem ataupun mencegah kerusakan sistem[9]. Salah satu tantangan adalah bagaimana menggunakan data log sistem operasional dan kinerja komputer paralel yang dihasilkan secara *real-time* oleh perangkat lunak Turbostat. Turbostat adalah alat pemantauan status prosesor yang menghasilkan data operasional dan kinerja prosesor dalam bentuk data log *time-series*[1]. Dengan menganalisis

data log memungkinkan kita untuk identifikasi anomali yang menunjukkan potensi kerusakan sistem[13]. Analisis data log sistem juga memungkinkan penyesuaian *workload* dan alokasi sumber daya secara *real-time*, yang berguna untuk meningkatkan efisiensi operasional dan pengoptimalan sistem. Namun, data ini sering tidak dapat dimanfaatkan sepenuhnya karena kurangnya *tools* atau metode khusus yang tersedia saat ini yang mampu menganalisis data log *streaming* secara efektif, baik dari segi komputasi maupun dari pemrosesan data secara *real-time*. Selain itu, analisis data *streaming* secara alami memiliki kelemahan dalam hal pemrosesannya karena membutuhkan komputasi yang besar[20]. Hal ini menjadikan salah satu tantangan dalam pemanfaatan analisis dengan data *streaming*. Maka dari itu, diperlukan teknik analitik yang tepat untuk hal ini.

*Functional Data Analysis* (FDA) memungkinkan analisis data *streaming*, atau yang selanjutnya disebut data kontinu dengan cara memanfaatkan fungsi kontinu untuk merepresentasikan data *streaming* secara *real-time*[16]. FDA memanfaatkan fungsi kontinu untuk representasikan data, yang memungkinkan kita untuk menggambarkan data kontinu dalam bentuk yang lebih mudah dipahami dan dianalisis. FDA juga memungkinkan kita untuk menggunakan teknik-teknik analisis yang lebih canggih, seperti regresi fungsional dan analisis variansi fungsional untuk memodelkan hubungan antara variabel-variabel dalam data kontinu[16]. Namun, metode FDA memiliki kelemahan dengan data yang memiliki dimensi tak terbatas, hal ini akan menimbulkan masalah komputasi yang tinggi.

Dalam mengatasi masalah-masalah tersebut, beberapa metode telah dikembangkan seperti *Multi Dimensionality Reduction* (MultiDR), metode ini berfokus pada pengurangan dimensi data kontinu secara *real-time* dengan mengurangi potensi kehilangan informasi[4]. Meskipun MultiDR dapat membantu dalam mengatasi tantangan dimensi tak terbatas, metode ini juga memiliki keterbatasan, yaitu dalam kompleksitas implementasi dan ketergantungan terhadap parameter, sehingga kurang dinamis untuk

diimplementasikan kuberbagai macam data[6]. Metode lainnya yaitu algoritma *Incremental and Progressive* untuk analisis *streaming* FDA, yang dinilai efektif dalam melakukan analisis terhadap data kontinu secara *real-time*, namun dengan biaya komputasi yang rendah[5]. Pada penelitian ini, algoritma *Incremental and Progressive* akan diimplementasikan untuk visualisasi dan analisis data *streaming* log sistem turbostat dalam konteks data operasional dan kinerja komputer paralel. Pemilihan algoritma *Incremental and Progressive* didasarkan pada kemampuannya yang dapat melakukan pemrosesan data *streaming* secara efisien dalam biaya komputasi. Maka dari itu, hal ini menjadi sebuah potensi yang cukup menjanjikan dalam hal analisis data operasional dan kinerja komputer paralel, karena dapat memungkinkan proses *monitoring* dan penanganan secara *real-time* dengan mempertimbangkan komputasi yang ringan.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah dijabarkan, ada beberapa rumusan masalah yaitu sebagai berikut:

1. Seberapa efektif penerapan algoritma *Incremental and Progressive* FDA dibandingkan algoritma FDA konvensional pada data log sistem hasil turbostat, untuk pemantauan dan analisis data operasional dan kinerja komputer paralel secara *real-time*.
2. Bagaimana mengimplementasikan dengan optimal metode algoritma *Incremental and Progressive* pada data log sistem hasil turbostat.

## **1.3. Tujuan**

Berdasarkan rumusan masalah yang sudah diidentifikasi, tujuan pada penelitian ini adalah :

1. Melakukan uji keefektifan dari metode algoritma *Incremental and Progressive* FDA dengan melakukan perbandingan dengan metode FDA konvensional.

2. Melakukan implementasi metode algoritma *Incremental and Progressive* yang optimal untuk data *streaming* log sistem turbostat.

#### **1.4. Batasan Masalah**

Terdapat beberapa batasan masalah dari penelitian yang dilakukan, yaitu sebagai berikut :

1. Data yang digunakan pada penelitian ini terdiri dari data operasional, kinerja, dan konsumsi energi yang saling terkait dan berbentuk *time series* dari log komputer paralel.
2. Data diambil menggunakan *tools* turbostat dengan melakukan uji coba *benchmark* dan dalam rentang waktu minimal satu jam.
3. Hasil penelitian berupa visualisasi dan faktor yang diukur adalah biaya komputasi, kelengkapan informasi dan visualisasi data.

#### **1.5. Metode Penelitian**

Terdapat beberapa metode penelitian pada penelitian ini, yaitu sebagai berikut:

##### **1. Studi literatur**

Metode ini dilakukan untuk mencari referensi dan informasi terkait topik penelitian yang dilakukan. Pada metode ini, pencarian referensi dan informasi dilakukan dengan pengambilan dari buku terkait dan melalui internet dari laman jurnal internasional. Dengan cara ini, studi literatur diharapkan dapat membantu dalam mendukung penelitian yang akan dilakukan.

##### **2. Implementasi metode**

Implementasi metode dilakukan dengan cara menerapkan metode dengan data terkait yang sebelumnya sudah dikumpulkan. Dari proses ini, nantinya hasil dari implementasi ini akan dilakukan proses analisis.

##### **3. Analisis**

Analisis dilakukan untuk meninjau dari hasil implementasi metode. Analisis dilakukan untuk mengukur kinerja dari metode yang digunakan,

dengan begitu langkah ini bertujuan untuk menilai hasil dari penelitian yang akan dilakukan.

### 1.6. Jadwal Pelaksanaan

Jadwal kegiatan yang dibuat berdasarkan rencana kegiatan dapat dilihat di Tabel 1.1.

Tabel 1.1 Jadwal Pelaksanaan Kegiatan Tugas Akhir

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan dan Eksplorasi Data						
3	Implementasi						
4	Analisis Hasil Penelitian						
5	Penyusunan Laporan/Buku TA						