

# BAB 1: PENDAHULUAN

## 1.1 Latar Belakang

Dengan berkembangnya teknologi informasi sekarang ini, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu bagian penting dalam kehidupan bermasyarakat saat ini ataupun di waktu mendatang. Tetapi kebutuhan informasi yang tinggi seringkali tidak diimbangi dengan pemberian informasi yang memadai, dalam penerapannya informasi masih harus di dalam dari data yang jumlahnya sangat besar. Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data tidak berbanding lurus dengan kemampuan untuk menganalisis, meringkas dan mengekstrak pengetahuan dari data itu sendiri. Metode tradisional untuk menganalisis data yang ada tidak dapat menangani data dalam jumlah besar.

Metode Gaussian Naive Bayes merupakan salah satu metode yang dapat dipakai untuk menjalankan proses klasifikasi *data mining*. Metode *Gaussian Naive Bayes* dipilih dikarenakan dataset yang akan di uji termasuk dataset kontinu dan menerapkan klasifikasi deterministik (tidak melibatkan nilai random/stokastik) sehingga waktu fase *training* lebih cepat, karena tidak perlu menggunakan siklus-siklus *training*[1].

Penggunaan GPU dapat mempercepat waktu proses dalam mengeksekusi program karena arsitekturnya yang *natively parallel*[2]. Selain itu Peningkatan performa yang terjadi tidak hanya berdasarkan kecepatan hardware GPU saja, tetapi faktor yang lebih penting adalah cara membuat kode program yang benar-benar bisa efektif berjalan pada GPU.

Pada karya tulis ini penulis akan menggunakan algoritma Gaussian naïve bayes untuk melakukan proses klasifikasi data mining pada GPU. Klasifikasi sendiri merupakan suatu proses menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu yang sudah ditentukan[3].

## 1.2 Perumusan Masalah

Perumusan masalah dalam tugas akhir ini dibagi kedalam beberapa poin, yaitu:

1. Bagaimana cara melakukan proses klasifikasi *data mining* pada GPU?
2. Bagaimana cara algoritma *Gaussian Naïve bayes* dapat melakukan pemrosesan data melalui GPU?
3. Bagaimana hasil *running time* yang terjadi pada GPU?

## 1.3 Batasan Masalah

Adapun batasan masalah pada tugas akhir ini adalah sebagai berikut:

1. Menggunakan *Gaussian Naïve bayes Algorithm*
2. Pemrosesan data difokuskan pada peningkatan *running time*.
3. *Hardware* yang digunakan harus menggunakan NVIDIA Graphic card.

## 1.4 Tujuan

Tujuan dari penulisan tugas akhir ini adalah:

1. Melakukan proses klasifikasi *data mining* yang berjalan pada GPU.
2. Melakukan pemrosesan data menggunakan *Gaussian Naive Bayes Algorithm* dengan menggunakan GPU
3. Menganalisis hasil *running time* yang dihasilkan dari proses klasifikasi *data mining* pada GPU.

## 1.5 Metodologi Penelitian

Adapun metodologi yang digunakan dalam pengerjaan tugas akhir ini adalah sebagai berikut:

### 1.5.1 Studi Literatur

Tahap ini dilakukan untuk memahami lebih dalam tentang pengujian kualitas sebuah perangkat lunak. Informasi didapatkan dari berbagai paper, jurnal maupun studi tidak formal dengan berdiskusi dengan orang-orang yang mempunyai minat dan latar belakang yang sama. Tentunya pendalaman literatur yang dilakukan selalu berhubungan dengan proses data mining yang dilakukan pada GPU dengan menggunakan metode *Gaussian*

*Naive Bayes*. Beberapa literatur berupa buku dan paper yang mendukung teori dan konsep tersebut sebagai berikut :

- a. *Data Mining*
- b. *A Brief Overview On Data Mining Survey*
- c. *Customer Classification And Prediction Based On Data Mining Technique*
- d. *Intrusion Detection Using Naive Bayes Classifier With Feature Reduction*
- e. *GPU-NB: A Fast CUDA- Based Implementation of Naive Bayes*
- f. *Parallel Data Mining on Graphic Processors*
- g. *Parallel and Distributed Data Mining*
- h. *A GPU Based Native Bayesian Algorithm for Document Classification*
- i. *The Optimality Naive Bayes*

### **1.5.2 Tahap Pencarian Data**

Pencarian Data dilakukan dengan mengunjungi situs dataset yang tersebar di dunia maya. Dataset yang digunakan diambil dari situs ucilab dan sudah dilakukan tes menggunakan *tools* weka dan mendapatkan akurasi 84%.

### **1.5.3 Preprocessing Data**

Data yang akan digunakan kemudian dilakukan tahapan preprocessing berupa data cleaning sehingga data yang tidak normal dapat dihilangkan dari dataset yang digunakan dan membantu mempercepat proses data mining.

### **1.5.4 Implementasi**

Implementasi dari tugas akhir ini adalah melakukan proses data mining menggunakan GPU dengan metode *Gaussian Naive Bayes* sehingga dihasilkan data berupa *initiation time, training time, testing time, total time, precision, recall* dan *accuracy* sehingga dapat dilakukan analisis terhadap proses *data mining* yang dilakukan.

