

Analisis Perbandingan CPU dan GPU (CUDA) Pada Klasifikasi Data Mining dengan Menggunakan Metode K-Nearest Neighbor Kernel Algorithm

Faris Muhammad¹, Ibnu Asror, S.T.,MT², Indra Lukmana Sardi. S.T.,MT³,

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹farismuhammad@students.telkomuniversity.ac.id, ²iasror@telkomuniversity.ac.id,

³indraluk@telkomuniversity.ac.id

Abstrak

Data mining merupakan proses semi-otomatis untuk mengeksplorasi data yang berjumlah besar gunanya untuk mendapatkan pola yang berguna. Data mining ini merupakan proses gabungan antar bidang-bidang terutama adalah machine learning, analisis statistik dan basis data. Data mining berusaha untuk menemukan kaidah dan pola dari data. Salah satu task yang penting dalam data mining adalah classification (klasifikasi). Klasifikasi ini dapat dideskripsikan sebagai berikut: terdiri dari data *input* yang disebut juga sebagai *training set* terdiri dari sejumlah *examples* (record) yang masing-masing memiliki sejumlah atribut atau disebut juga fitur. Adapun tujuan klasifikasi ini adalah untuk menganalisa data *input* dan mengembangkan sebuah model yang akurat untuk setiap kelas berdasarkan beberapa *variabel prediktor*. Untuk menghasilkan informasi saat melakukan proses data *mining* kendala yang dihadapi adalah banyaknya jumlah data sehingga proses yang dilakukan oleh CPU akan berjalan sangat lambat apabila dirasakan. Untuk menanggulangi masalah ini maka proses data mining menggunakan GPU menjadi salah satu solusi dalam menangani *running time* yang lambat dan akurasi yang kurang baik. Melalui tugas akhir ini penulis akan mencoba menganalisis sebuah algoritma KNN Kernel, Metode ini merupakan perkembangan dari metode KNN *Standard*. Dimana pada metode KNN *Standard* proses klasifikasi dilakukan dengan melihat sejumlah *k* tetangga terdekat, dan akan diklasifikasikan berdasarkan jumlah kelas terbanyak pada sejumlah *k* tetangga terdekatnya. *Classifier* tersebut diuji menggunakan 3 fungsi Kernel. Hasil yang didapat dari percobaan penulis yaitu pada pembagian *5 fold* total waktu CPU1: 1,68 s, CPU2: 15,63 s, GPU1: 12,29 s, GPU2: 4,61 s. dan pada pembagian *10 fold* total waktu CPU1: 1,53 s, CPU2: 15,27 s, GPU1: 12,05 s, GPU2: 4,55. Akurasi yang didapatkan pada pembagian *5 fold* 63,87% dan pembagian *10 fold* 64,30% pada semua perangkat.

Kata Kunci : data mining, klasifikasi, CPU, GPU, KNN Kernel

Abstract

Data mining is a semi-automatic process for exploring and analyzing large amounts of data to get useful patterns. Data mining is a joint process between fields, especially machine learning, statistical analysis and database. Data mining tries to find the rules and patterns of data. One important task in data mining is classification (classification). This classification can be described as follows: consists of input data which is also called training set consisting of a number of examples (records) which each have a number of attributes or also called features. The purpose of this classification is to analyze input data and develop an accurate model for each class based on several predictor variables. To produce information when doing data mining process, the obstacles faced are the large amount of data so that the process carried out by the CPU will run very slowly when felt. To overcome this problem, the data mining process uses GPU to be one of the solutions in handling slow running time and poor accuracy. Through this final project the author will try to analyze a KNN Kernel algorithm, this method is a development of the KNN Standard method. Where in the KNN Standard method the classification process is carried out by looking at a number of the closest neighbors, and will be classified based on the number of classes in the number of the closest neighbors. The classifier is tested using 3 Kernel functions. The results obtained from the authors' experiments are that the division of 5 fold total CPU time1: 1.68 s, CPU2: 15.63 s, GPU1: 12.29 s, GPU2: 4.61 s. and in dividing the 10 fold total CPU time1: 1.53 s, CPU2: 15.27 s, GPU1: 12.05 s, GPU2: 4.55. Accuracy obtained at 5 fold division is 63.87% and division of 10 fold is 64.30% on all devices.

Keywords: data mining, classification, CPU, GPU, KNN Kernel

1. Pendahuluan

1.1 Latar Belakang

Seiring dengan perkembangan teknologi dalam hal pengumpulan data dan penyimpanan data dapat menyebabkan tumpukan data yang banyak. Dengan adanya kumpulan data yang banyak, maka timbulah suatu kebutuhan untuk bisa memanfaatkan data tersebut. Pemanfaatan data tersebut tentunya bertujuan untuk mendapatkan informasi penting dari pola-pola data yang terbentuk. Proses untuk mendapatkan informasi atau pola-pola berharga dari sekumpulan data tersebut lah dinamakan Data mining [1]. Klasifikasi merupakan suatu metode dari data mining. Ini merupakan metode prediktif yang melakukan pembelajaran terhadap data-data yang sudah