

1. Pendahuluan

1.1. Latar Belakang

Perkembangan ilmu pengetahuan dan teknologi informasi yang sangat cepat memberikan kemudahan dalam pencarian informasi. Salah satunya dapat dilihat dari semakin banyaknya jumlah *website* yang ada pada dunia maya. Pemanfaatan dalam penggunaan *website* banyak terdapat dalam berbagai bidang, salah satunya adalah *e-commerce*. *E-commerce* merupakan istilah yang digunakan dalam pembelian, penjualan dan proses pemasaran barang yang memanfaatkan teknologi yang berkembang pada saat ini yaitu internet. *Website e-commerce* yang digunakan pada tugas akhir ini adalah *tuneeca online store* yang dibangun oleh perusahaan Bippo Indonesia. Website ini melakukan proses transaksi jual beli dan pemasaran pada barang berupa busana muslim modern. Website e-commerce ini dipilih karena memiliki *traffic* yang tinggi dan user dapat bebas mengakses konten yang terdapat pada website tersebut .

Interaksi yang terjadi selama pengaksesan *website* tersimpan secara otomatis pada *web server* dalam bentuk *web server log file*. Setiap kali terdapat interaksi pada website, *web server* menyimpannya dalam satu baris *log*. Data yang disimpan pada *web server log* umumnya terdiri dari informasi mengenai *user* yang mengakses serta interaksi apa yang dilakukan pada *website*. Analisa yang dilakukan pada *web server log* telah banyak digunakan untuk melihat aktifitas pengunjung web, seperti menentukan jumlah akses terhadap halaman web, jumlah user yang mengakses web dan waktu mengunjungi web, umpan balik dari perilaku user terdapat dalam web dan masih banyak lagi.

Web usage mining merupakan salah satu kategori dari *web mining* berdasarkan jenis data primer yang bertujuan untuk menangkap dan memodelkan pola perilaku dan profil dari pengunjung website. *Web usage mining* mengaplikasikan berbagai teknik *data mining* seperti *classification*, *clustering*, dan *association rules*. Teknik permodelan yang dapat digunakan dalam mengolah pola perilaku user adalah *clustering dan classification*. *Clustering* adalah pengelompokan objek-objek yang berkaitan atau memiliki kemiripan. *Classification* adalah suatu proses menemukan model yang menjelaskan dan membedakan kelas data atau konsep, dengan tujuan untuk dapat memprediksi kelas objek dari label kelas yang belum diketahui. Model yang didapatkan dari hasil *classification* didasarkan pada analisis dari satu set data pelatihan [14].

Algoritma K-Means digunakan sebagai metode *clustering* untuk membentuk kelompok user yang memiliki kemiripan dalam mengakses halaman *website* ke dalam cluster dengan sangat cepat. Algoritma K-Means selain digunakan untuk proses clustering dalam pelabelan data juga merupakan *semi-supervised learning* sehingga dapat melakukan pengklasifikasian terhadap data yang belum memiliki label [1]. Tetapi algoritma ini mengelompokkan datanya tanpa ada pelatihan terlebih dahulu yang menyebabkan hasil akurasi rendah, sehingga untuk mengklasifikasikan data akan digunakan algoritma LVQ. *Learning Vector Quantization (LVQ)* adalah sebuah metode pengelompokan dimana setiap unit output mempresentasikan sebuah kelas [3]. Model pembelajaran LVQ dilatih secara signifikan sehingga lebih cepat dibandingkan algoritma lainnya seperti *Back Propagation Neural Network*. Algoritma ini akan digunakan pada tahap *classification* untuk mengetahui tingkat ketertarikan pengguna dalam menggunakan website *Tuneeca*. Dengan adanya metode klasifikasi ini dapat dilakukan prediksi ke user yang belum memiliki clusternya, sehingga dapat dimanfaatkan untuk mendapatkan informasi hubungan antara user dengan halaman website pada data yang terdapat pada bulan selanjutnya.

1.2. Perumusan Masalah

Berdasarkan uraian latar belakang, maka permasalahan yang dapat dirumuskan pada penelitian ini adalah:

1. Bagaimana menerapkan algoritma K-Means dalam proses *clustering* pada *Web Usage Mining*?
2. Bagaimana menerapkan algoritma LVQ serta performansinya dalam menerapkan teknik *classification* pada *Web Usage Mining*?
3. Bagaimana hasil analisis dari prediksi user *tuneeca* dengan menggunakan metode *classification LVQ*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah :

1. Menerapkan algoritma *K-Means* dalam proses *clustering* pada *Web Usage Mining*.
2. Menerapkan algoritma LVQ dan mengetahui performansinya dalam menerapkan teknik *classification* pada *Web Usage Mining*
3. Melakukan analisis dari hasil prediksi user *tuneeca* dengan menggunakan metode klasifikasi *LVQ (Learning Vector Qualization)* dalam mengelompokkan user ke dalam cluster

1.4. Batasan Masalah

Dalam tugas akhir ini, permasalahan dibatasi dalam beberapa hal yaitu:

1. Data yang digunakan adalah web server log dari situs www.tuneecca.com/
2. Data yang digunakan merupakan web server log pada 1 bulan antara 21 Juli 2014 - 26 Agustus 2014
3. Pola navigasi user yang terbentuk berdasarkan hubungan antar halaman yang ada pada *tuneecca* yang terdiri dari 47 halaman.
4. Pengaruh parameter yang dianalisis pada algoritma K-Means adalah jumlah kluster (k)
5. Pengaruh parameter kualitas metode klasifikasi LVQ adalah *learning rate*, dan maksimum epoch.
6. Proses *preprocessing* yang digunakan adalah metode *Session Identification* standart untuk *Web Usage Mining*
7. Sistem yang dibuat masih bersifat Offline.

1.5. Metodologi Penyelesaian Masalah

Metode penelitian yang digunakan dalam penelitian ini terdiri atas delapan tahap, yaitu

:

1. Studi literatur
Studi literatur yang dilakukan bertujuan untuk proses pembuatan laporan tugas akhir dengan mengacu referensi-referensi yang berkaitan dengan *Web Usage Mining*, K-Means dan LVQ beserta dengan tools yang akan digunakan dalam membangun sistem.
2. Pengumpulan data
Mengumpulkan data web log server yang digunakan untuk bahan analisis *web usage mining*.
3. Analisis dan Perancangan Sistem
Membuat spesifikasi kebutuhan sistem dan merancang proses *preprocessing*.
4. Implementasi dan pembangunan sistem
Sistem yang akan dibangun ini akan diimplementasikan dengan menggunakan bahasa pemrograman java serta tools berupa Netbeans IDE 7.1 dan MySQL.
5. Pengujian Sistem
Pengujian sistem dilakukan dengan melihat hasil keluaran sistem berupa kluster modul-modul yang ada pada web *tuneecca* yang merepresentasikan pola akses user yang

memiliki kemiripan serta mengamati performansi dari sistem dan akurasi klasifikasi user ke dalam klaster modul-modul tersebut.

6. Penyusunan Laporan

Penyusunan laporan tugas akhir merupakan tahap terakhir dalam pembangunan sistem yang dilakukan untuk mendokumentasikan penelitian yang telah dilakukan supaya dapat bermanfaat bagi berbagai pihak.

1.6. Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika pembahasan sebagai berikut :

Bab 1 Pendahuluan

Bab ini memaparkan latar belakang dilakukannya penelitian, tujuan yang ingin dicapai dari penelitian ini, batasan masalah, perumusan masalah yang akan dibahas, metode penyelesaian masalah, dan sistematika penulisan.

Bab 2 Tinjauan Pustaka

Bab ini memuat penjelasan mengenai teori-teori yang mendukung dan melandasi penulisan tugas akhir, yaitu tentang metode algoritma K-Means, LVQ serta teori mengenai *web usage mining*.

Bab 3 Perancangan Sistem

Bab ini berisi mengenai analisis kebutuhan perangkat lunak maupun perancangan sistem secara global dengan menggunakan diagram alir, diagram konteks, serta data flow diagram.

Bab 4 Implementasi Dan Pengujian Sistem

Bab ini berisi tentang implementasi dari sistem yang dibangun dengan menggunakan bahasa pemrograman Java. Bab ini juga berisikan mengenai cara kerja dan fungsionalitas dari sistem yang akan dibangun, serta cara dan hasil pengujian yang dilakukan dari sistem tersebut.

Bab 5 Kesimpulan Dan Saran

Bab ini berisikan tentang kesimpulan dari hasil penelitian tugas akhir ini dan saran-saran untuk pengembangan lebih lanjut.