

BAB 1 Pendahuluan

1.1. Latar Belakang

Saat ini, pertumbuhan data *web* di internet semakin besar dan perkembangan *user* dengan *behaviour* masing masing juga semakin beragam. *Website* dengan konten yang serupa banyak ditemui *user* di internet. Hal tersebut menjadikan tantangan tersendiri bagi para pengembang *website* untuk merumuskan bagaimana menarik perhatian *user* saat melakukan aktivitas di *website* mereka. Memahami karakteristik dan preferensi *user* atau yang disebut sebagai *personalization*. *Personalization* menjadi satu hal yang sangat penting dalam meningkatkan *experience user* saat melakukan aktivitas di suatu *website*^[1]. *Web page prediction* merupakan suatu *classification problem* yang mampu memprediksi *next page* yang akan dikunjungi *user* berdasarkan *knowledge* dari *page-page* yang dikunjungi sebelumnya. *Web page prediction* dapat diimplementasikan secara praktis dalam aplikasi yang digunakan di bidang industri seperti *search engine*, *caching system*, *recommendation system*, dan *wireless application* ^[2]. Hasil dari suatu *page prediction* dapat digunakan sebagai *personalization* sebuah *web*, pengurangan waktu respon *server* dengan strategi *prefetching* dan *caching* yang sesuai. Hal ini dapat digunakan sebagai pedoman untuk menambahkan *recommendation system*, meningkatkan *design* dari aplikasi suatu *web*, *e-commerce* untuk handle isu bisnis yang spesifik seperti *customer attraction*, *customer retention*, *cross sales* dan *customer departure* ^[3].

Dalam *web usage mining*, terdapat analisis informasi dari *web pages* yang pernah dikunjungi *user*. Data tersebut tersimpan di *log files server* yang nantinya digunakan dalam menemukan pola pola ketertarikan yang sangat berguna ^[4]. Tiga *task* yang secara umum dilakukan dalam *web usage mining* ialah *data preprocessing*, *pattern discovery*, dan *pattern analysis*^[5]. Namun dalam memprediksi *next page user*, tantangan akan dihadapi baik dalam tahap *preprocessing* maupun *prediction*. Dalam *preprocessing*, beberapa tantangan yang dihadapi ialah besarnya ukuran data yang diproses dan identifikasi *session user* dalam *data log*. Sementara di tahap *prediction*, *long time prediction* dan *low prediction accuracy* serta *memory limitation* akan menjadi tantangan dalam menghasilkan *next page prediction* yang sesuai dengan *user*^[2].

Selama ini *Markov model* telah dikenal dengan efisiensi dan performanya dalam menghasilkan prediksi, hanya saja semakin rendah *order* dari *Markov model* berakibat pada menurunnya akurasi prediksi sementara semakin tinggi *order Markov model* akan berakibat pada kompleksitas program yang bertambah ^[6]. Namun *Markov model* pun masih memiliki kemungkinan tidak dapat menghasilkan prediksi atau menghasilkan prediksi yang tidak termasuk dalam kelas mayoritas^[7].

Oleh karena itu, dengan motivasi menghasilkan suatu model prediksi yang memiliki akurasi prediksi yang tinggi, serta mencari alternatif ketika *Markov model* tidak mampu menghasilkan prediksi karena adanya *zero probability*, maka dalam pengerjaan Tugas Akhir ini, sebuah model yang mampu memprediksi *next page* yang akan dikunjungi *user* akan diimplementasikan dengan pendekatan kombinasi teknik *Markov model* serta *Association rule mining (ARM)* dengan *data set* yang

digunakan ialah *data log server* dari salah satu situs *e-commerce* yang ada di Indonesia, yakni Bukalapak (www.bukalapak.com). *Association rule mining*(ARM) dipilih karena teknik ini akan membaca *history access page user* dan *state session* diperiksa lebih banyak untuk menghasilkan *rule* yang lebih general. *Rule* yang didapat dari ARM inilah yang akan digunakan untuk menghasilkan prediksi ketika *Markov model* tidak menghasilkan prediksi.

1.2. Perumusan Masalah

Berdasarkan pada latar belakang diatas, rumusan masalah yang akan diuraikan dan diteliti adalah sebagai berikut:

1. Bagaimana memprediksi *next page* yang akan dikunjungi *user* berdasarkan data pada *log server*
2. Bagaimana modifikasi dengan penggabungan *Markov model* dan ARM diterapkan pada *web page prediction*
3. Bagaimana pengaruh penggunaan kombinasi *Markov model* dan ARM terhadap performa *web page prediction*
4. Bagaimana melakukan *preprocessing* terhadap data *log server* yang membantu *web page prediction* dalam menghasilkan prediksi yang optimal

Dalam penelitian Tugas Akhir ini, terdapat beberapa batasan masalah dalam pembahasan dan pengerjaannya, diantaranya :

1. Dataset yang digunakan berupa data *log server* untuk *public pageview* tentang produk produk yang dijual di *website* Bukalapak. Data *log server* yang digunakan ini ialah data *log* pada tanggal 10, 12, 14, dan 15 Februari 2016 dan berupa data .csv
2. Penentuan prediksi *next page* yang akan dikunjungi *user* hanya didasarkan pada satu *page* yang dikunjungi sebelumnya
3. *Page* yang dimaksud dalam *web page prediction* ialah *page* yang merupakan sub terkecil dari *page* kategori belanja pada situs Bukalapak bukan *page* yang merupakan produk pada katalog belanja di situs Bukalapak
4. Sistem yang dibuat masih bersifat *offline* dalam artian tidak langsung diembed ke situs *e-commerce* Bukalapak

1.3. Tujuan

Adapun tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. Mengimplementasikan suatu model yang mampu memprediksi *next page* yang akan dikunjungi *user* berdasarkan data pada *log server*
2. Mengetahui cara penerapan modifikasi dengan penggabungan *Markov model* dan ARM pada *web page prediction*
3. Mengetahui pengaruh penggunaan kombinasi *Markov model* dan ARM terhadap performa *web page prediction*
4. Mengetahui cara melakukan *preprocessing data log server* yang membantu *web page prediction* dalam menghasilkan prediksi yang optimal

1.4. Metodologi Penyelesaian Masalah

Pengerjaan tugas akhir ini dilakukan dengan beberapa metodologi penyelesaian masalah sebagai berikut:

1. Kajian Pustaka
Pada tahap ini, akan dilakukan pencarian dan pengumpulan berbagai macam referensi tentang *Web Usage mining*, *Web Page Prediction*, *K-th order Markov model*, *Association rule mining*, *Apriori Algorithm* yang terdapat pada jurnal, paper, atau e-book dan mempelajari materi tersebut.
2. Pengumpulan Data
Pada tahap ini bertujuan untuk lebih mendalami materi yang telah didapatkan pada tahap sebelumnya yaitu tahap Kajian pustaka. Pada tahap ini juga akan dilakukan analisis permasalahan yang dijadikan topik penelitian dan analisa data yang akan digunakan pada penelitian ini. Hasil dari tahap ini yaitu telah ditentukan metode yang akan digunakan dalam penyelesaian masalah yang diangkat serta mendapatkan data yang akan digunakan pada penelitian ini, data yang digunakan adalah data log server untuk public *pageview* tentang produk produk yang dijual di Bukalapak dalam bentuk file .csv.
3. Analisis dan Perancangan Sistem
Pada tahap ini, akan dilakukan perancangan sistem dengan membuat flowchart dari setiap tahapan yang dilakukan, dengan membuat perancangan sistem ini akan mudah memahami alur yang terjadi didalam sistem yang akan dibangun.
4. Implementasi Sistem
Pada tahap ini, akan dilakukan implementasi dari setiap alur yang terdapat pada perancangan sistem yang telah dibuat pada tahap sebelumnya, implementasi pada Tugas akhir ini dibangun menggunakan bahasa pemrograman Python dengan tools PyCharm Community Edition 2016.2.1
5. Analisis dan Pengujian
Pada tahap ini, akan dilakukan pengujian terhadap sistem yang telah dibangun untuk mengetahui hasil implementasi yang telah dilakukan serta menganalisis dari hasil pengujian yang dilakukan .
6. Pembuatan Laporan Tugas Akhir
Pada tahap ini, akan dilakukan penyusunan laporan dan pengumpulan dokumentasi berdasarkan hasil dari penelitian.

1.5. Sistematika Penulisan

Penulisan tugas akhir ini akan dibagi menjadi lima bab yaitu:

1. Bab 1 : Pendahuluan
Pada bab ini akan berisikan latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah, dan sistematika penulisan
2. Bab 2 : Tinjauan Pustaka
Bab ini memuat teori-teori yang terkait dengan pengerjaan tugas akhir ini. Teori yang dimuat ialah teori mengenai *Web Usage mining*, *Web Page*

Prediction, K-th order Markov model, Association rule mining, Apriori Algorithm.

3. Bab 3 : Perancangan Sistem

Pada bab tiga ini akan dibahas perancangan sistem yang akan digunakan untuk membuat *prediction model* terhadap *page* yang akan dikunjungi *user* di situs Bukalapak menggunakan modifikasi penggabungan *Markov model* dan *Association rule mining*

4. Bab 4 : Implementasi dan Analisa Sistem

Bab ini akan membahas dokumentasi dan analisis dari pengujian *prediction model* terhadap *page* yang akan dikunjungi *user* di situs Bukalapak menggunakan modifikasi penggabungan *Markov model* dan *Association rule mining* dengan menggunakan dataset yang ada.

5. Bab 5 : Kesimpulan dan Saran

Bab ini akan memuat kesimpulan beserta saran dari hasil pengerjaan tugas akhir ini.