

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Seiring dengan perkembangan zaman, kemajuan internet berkembang dengan sangat pesat, karenanya setiap organisasi mempromosikan dirinya di internet. Tolak ukur keberhasilan dari sebuah website adalah tingginya jumlah pengunjung, untuk menjaga pengunjung agar tetap mengakses sebuah website tersebut, dibutuhkan sistem prediksi apakah pengguna akan kembali mengakses website tersebut atau tidak. Penting untuk diingat karena deteksi kesamaan dibangun untuk mendukung prediksi konsistensi akses web, maka deteksi kesamaan yang akan diobservasi adalah website ke website dan website ke software.

Sistem prediksi tersebut dapat didasarkan dari log user behavior, sebagai contoh ketika seseorang yang tadinya mengakses website tentang musik dan pada seminggu terakhir sudah mulai mengakses website tentang game, ada kemungkinan jika dia tidak akan mengakses website musik kembali, karena ini adalah masa transisi ketertarikan seseorang dari musik ke game.

Untuk menentukan apakah dua buah website atau software memiliki genre yang sama, misalnya musik dan game yang disebutkan pada contoh sebelumnya diperlukan sebuah mekanisme untuk menentukan sebuah kesamaan konteks. Pada penelitian ini sistem deteksi tersebut akan dibangun dengan menggunakan teknik k-means clustering dan cosine similarity.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang yang ada, maka dapat diambil rumusan masalah menjadi sebagai berikut:

1. Bagaimana cara mendefinisikan kesamaan topik dari kedua buah website?
2. Apakah deteksi similaritas dua website efektif?

### 1.3 Tujuan Pembahasan

Berdasarkan rumusan masalah yang ada, maka tujuan pembahasan dari tugas akhir ini adalah sebagai berikut:

1. Kesamaan dua topik ditentukan dari *clustering* dan *cosine similarity*.

2. Menguji deteksi kesamaan yang telah dibuat berdasarkan relevansi berdasarkan kata yang banyak muncul.

#### **1.4 Ruang Lingkup**

Ruang lingkup pembahasan dalam penulisan tugas akhir ini dibatasi dengan data-data yang diberikan oleh Guizhou University. Data yang dipakai terdiri dari nama *browser* dan alamat *web* yang digunakan oleh *user*. Data yang diberikan terdiri dari 980 pengguna. Untuk konten yang diambil pada website diasumsikan gambar tidak berpengaruh terhadap konten.

#### **1.5 Sumber Data**

Pada penulisan tugas akhir ini, sumber data yang digunakan adalah sebagai berikut:

1. Sumber data primer : Data log *user behavior* yang diberikan oleh Guizhou University.
2. Sumber data sekunder : Data sekunder didapatkan dari jurnal, buku-buku, dan halaman *website* yang berhubungan dengan pengelompokan data.

#### **1.6 Sistematika Penyajian**

Sistematika penulisan yang digunakan dalam laporan ini adalah sebagai berikut:

##### **BAB I. PENDAHULUAN**

Bab ini berisi mengenai latar belakang, rumusan masalah, tujuan pembahasan, ruang lingkup, sumber data, dan sistematika penyajian dalam pembuatan tugas akhir ini.

##### **BAB II. KAJIAN TEORI**

Bab ini berisi mengenai teori yang berhubungan dengan penyelesaian tugas akhir ini.

##### **BAB III. ANALISIS DAN RANCANGAN SISTEM**

Bab ini berisi mengenai analisis metodologi yang berupa *pseudocode* yang digunakan dalam pembuatan tugas akhir ini.

##### **BAB IV. IMPLEMENTASI**

Bab ini berisi mengenai implementasi *pseudocode* dalam pembuatan tugas akhir ini.

## BAB V. PENGUJIAN

Bab ini berisi hasil pengujian sistem dalam pembuatan tugas akhir ini.

## BAB VI. SIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran dari pembahasan dan perancangan serta analisa pengujian sistem yang dibuat dalam pembuatan tugas akhir ini.

