

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam *deep learning*, data memiliki peran yang sangat penting dalam menginterpretasikan sebuah data sebagai komposisi objek, seseorang akan menyukai untuk secara otomatis mengidentifikasi apa yang objek yang di dapatkan dan mendeteksi objek [2]. Data di internet memiliki jumlah yang sangat banyak, tetapi hanya sebagian kecil dari informasi tersebut yang dikumpulkan dengan metadata [4]. *Scraping data* adalah teknik di mana data dari situs web HTML (*Hypertext Markup Language*) diekstraksi dan disimpan ke dalam basis data secara otomatis. *Data* yang dihasilkan kemudian digunakan untuk pengambilan atau analisis [3].

Kuantitas dan kuantitas *data* biasanya memiliki pengaruh besar dalam hasil penelitian dan membuat pengumpulan *data* sebagai peran penting dalam proses penelitian. Pengumpulan *data* dari banyak tautan situs yang disediakan memiliki format yang berbeda [11]. Banyak *search engine* seperti google chrome, firefox dan lainnya mempunyai program untuk berinteraksi dengan halaman dengan mengklik tautan dalam melihat seluruh isi dari semua *data*. *Webdriver Selenium* dapat mengotomatiskan aplikasi tautan situs untuk tujuan pengujian, selain itu memudahkan penggunaannya, dan ketersediaan driver yang dikonfigurasi *selenium* untuk berbagai mesin pencari situs tautan [8].

Pengambilan *data* menggunakan *selenium* menyediakan banyak operasi untuk pengujian situs aplikasi dan juga memiliki kemampuan mendukung halaman situs yang bersifat dinamis yaitu untuk mendukung halaman yang elemennya dapat berubah tanpa memuat ulang halaman itu sendiri [7]. *Data* gambar yang akan di peroleh dari hasil *scraping* di *Google Images* adalah tautan aslinya, juga seluruh gambar yang terdapat di *Google Images*, dan ada pula kombinasi dari hasil relasi yang di dapatkan dari hasil *keyword* yang dimasukan.

1.2 Rumusan Masalah

Adapun perumusan masalah dari latar belakang diatas adalah:

1. Sistem pencarian seperti apa yang cocok untuk mengambil seluruh tautan data atau gambar yang ada di *google images*.
2. Bagaimana rancangan sistem relasi atau kombinasi pada pengambilan tautan gambar.

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Membuat system *engine* untuk mendapatkan seluruh tautan asli yang dicari dengan menggunakan library selenium.
2. Mengambil seluruh data atau gambar dengan seluruh relasi yang sudah di kombinasikan dari hasil keyword yang di dapatkan.

1.4 Batasan Masalah

Adapun batasan masalah dari penelitian tugas akhir ini adalah:

1. *Search Engine* yang digunakan untuk mencari tautan gambar menggunakan chrome.
2. Pengambilan tautan gambar yang di dapatkan adalah berupa alamat atuatan aslinya.
3. Pencarian dan pengambilan tautan gambar meliputi relasi kategori yang sudah di kombinasikan menggunakan *library itertools* dan *selenium*.
4. Prediksi ukuran besaran data menggunakan regresi linear.
5. Kombinasi relasi kategori hanya terdapat pada kategori bagian awalan.
6. Kombinasi dari *itertools* hanya dilakukan sebanyak 4.
7. Hasil tautan gambar dari *scraping data* berupa bentuk teks.

1.5 Metodologi Penelitian

Metodologi penelitian yang digunakan paga Tugas Akhir ini adalah:

1. Studi Literatur

Penulis melakukan studi literature dengan mengumpulkan dan mempelajari teori serta konsep-konsep dari sumber penunjang materi *Google*

Images, Selenium Webdriver, dan Scraping. Sumber literature itu meliputi *paper, website, blog, dan jurnal*. Kegiatan ini agar penulis mendapatkan pemahaman yang cukup pembuatan Tugas Akhir.

2. Analisis Kebutuhan

Setelah dilakukan studi literature, akan dilakukan analisis kebutuhan penelitian yang akan digunakan untuk mendapatkan tautan gambar dan relasi kategori yang dibutuhkan untuk analisis penelitian.

3. Perancangan Aplikasi

Pada tahap ini membuat perancangan program, pemodelan data, dan regresi linear yang kemudian diterapkan menjadi *engine scraping data*.

4. Pengujian Aplikasi

Program yang dibuat kemudian di uji dengan pengujian *white box*, pengujian *stabilitas*, dan pengujian *beta*.

5. Diskusi Ilmiah

Proses diskusi ilmiah dilakukan dengan dosen Pembimbing untuk memperbaiki kesalahan yang didapat pada *engine* yang telah dibuat.

6. Pembuatan Laporan

Pada proses ini dilakukan pembuatan proses dokumentasi laporan Tugas Akhir yang akan digunakan untuk sidang Tugas Akhir.

1.6 Sistematika Penulisan

Penulisan tugas akhir ini dibagi dalam beberapa bagian. Tiap-tiap bagian menjelaskan langkah demi langkah dalam pengerjaan tugas akhir ini. Berikut adalah bagian tersebut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan latar belakang penelitian, rumusan masalah tujuan, dan batasan masalah, dan metodologi penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang tinjauan teori dan sumber-sumber terkait yang digunakan dalam penelitian yang dilakukan.

BAB III ANALISIS DAN PERANCANGAN

Bab ini menjelaskan tentang proses analisis pada sistem serta desain dan perancangan aplikasi system sesuai tinjauan teori yang dipakai pada penelitian tugas akhir ini.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas proses pengujian yang dilakukan pada aplikasi, dan hasil pengujian dianalisis agar dapat dilakukannya penarikan kesimpulan dalam penelitian ini.

BAB V KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan dari hasil pengujian aplikasi yang dibuat dan saran yang membuat hasil penelitian lebih baik untuk kedepannya.