

## **BAB II**

### **TINJAUAN PUSTAKAN DAN DASAR TEORI**

#### **2.1 Tinjauan Pustaka**

Penelitian dilakukan oleh Obet Edom Martinus Fransisco (2019) di STMIK AKAKOM Yogyakarta, penelitian berfokus pada pengumpulan informasi mengenai promo dari beberapa toko online. Mengimplementasikan dasar web scraping dengan menggunakan simple HTML DOM untuk melakukan setiap parsing data melalui pemrograman php.

Penelitian kedua tentang implementasi teknik web scraping pernah dilakukan oleh Chunmei Zheng, Guomei Hei, dan Zuojie Peng (2015) di China University of Geoscience, Beijing. Mereka berfokus tentang studi teknologi ekstraksi informasi web yang menyediakan informasi lowongan pekerjaan untuk lulusan baru universitas secara akurat, menggunakan teknik web scraping berdasarkan struktur DOM melalui package Beautiful Soup dari Python.

Penelitian ketiga berfokus tentang implementasi teknik web scraping dengan klasifikasi naïve bayes pernah dilakukan oleh Dave Juliano, Agustinus Noertjahya, Anita Nathania Purbowo (2020) di Universitas Kristen Petra Surabaya. Mereka bertiga mengumpulkan informasi dari situs penyedia informasi promo menggunakan package Beautiful Soup dari Python berdasarkan struktur DOM, lalu untuk mendapatkan informasi promo mengenai makanan dilakukan proses klasifikasi menggunakan naïve bayes untuk mengklasifikasi promo yang sesuai dengan kategori makanan.

Penelitian keempat dilakukan oleh Dhita Deviacita Ayani, Helen Sasty Pratiwi, dan Hafiz Muhandi (2019) di Universitas Tanjungpura, Kalimantan Barat. Penelitian ini membahas mengenai teknik web scraping dengan parsing HTML DOM menggunakan bahasa pemrograman php untuk mengumpulkan data produk dari sebuah situs marketplace berdasarkan kata kunci. Penelitian ini ditujukan untuk membantu pengguna untuk mendapatkan produk yang diinginkan dengan hasil penjualan yang terbaik dan memiliki perbandingan harga dari berbagai situs.

Penelitian kelima memiliki fokus tentang web scraping pernah dilakukan oleh Dana Febri Setiawan, Tristiyanto, Astria Hijriani (2020) di Universitas Lampung, Bandar Lampung. Menggunakan bahasa pemrograman php untuk melakukan parsing HTML demi mendapatkan informasi mengenai data produk dari situs marketplace yang mempermudah pengguna untuk mencari sebuah produk dan deskripsi produk dengan cepat dan mudah.

Dari kelima referensi di atas, ada perbedaan penerapan dari web scraping yang pernah dibuat dan yang belum pernah dibuat yaitu belum ada penerapan web scraping untuk data quick count pilpres per provinsi Indonesia.

Bisa dilihat pada table 2.1 untuk perbandingan penelitian yang sudah pernah dilakukan.

Tabel 2.1 Perbandingan Metode Penelitian

No	Peneliti	Objek	Tujuan Penelitian	Teknik yang Digunakan	Informasi yang Dihasilkan
1	Fransisco (2019)	Penyedia Promo toko online	Implementasi dasar <i>web scraping</i> untuk mengumpulkan informasi promo pada toko online	Teknik <i>web scraping</i>	Website Daftar promo pada 10 toko online
2	Chunmei, Guomei, dan Zuojie (2015)	Penyedia informasi lowongan pekerjaan	Implementasi <i>web scraping</i> untuk membantu mahasiswa lulusan baru yang mencari lowongan pekerjaan	Teknik <i>web scraping</i>	Website pencarian lowongan pekerjaan berdasarkan kata kunci
3	Dave, Agustinus, dan Anita (2019)	Situs penyedia promo	Implementasi <i>web scraping</i> untuk membantu mengumpulkan informasi promo makanan	Teknik <i>web scraping</i> dan klasifikasi <i>Naïve Bayes</i>	Aplikasi penyedia promo makanan dari beberapa website penyedia promo
4	Dhita, Helen, dan Hafiz (2020)	Data produk <i>marketplace</i>	Implementasi <i>web scraping</i> untuk pengumpulan data produk <i>marketplace</i>	Teknik <i>web scraping</i>	Website pencarian produk dari <i>marketplace</i>
5	Dana, Tristiyanto, dan Astria (2020)	Data produk <i>marketplace</i>	Implementasi <i>web scraping</i> untuk membantu mendapatkan deskripsi produk	Teknik <i>web scraping</i>	Website penyedia data dan deskripsi produk dari berbagai situs <i>marketplace</i>
6	Irfan (2021)	Data <i>quick count</i> pilpres	Implementasi <i>web scraping</i> untuk mengumpulkan data <i>quick count</i> pilpres	Teknik <i>web scraping</i>	Database dan API yang menampilkan data <i>quick count</i> pilpres per provinsi

## **2.2 Dasar Teori**

### **2.2.1 Python**

Python Trisanti (2002) adalah bahasa pemrograman tingkat tinggi berbasis skrip yang pertama kali dikembangkan oleh Guido van Rossum pada tahun 1990 di CWI, Ansterdam sebagai penerus dari bahasa pemrograman ABC. Pada tahun 1995, Guido pindah ke CNRI dan tetap melanjutkan pengembangan Python sampai dengan versi 1.6. Tahun 2000 Guido dan para pengembang inti pindah ke sebuah perusahaan komersil bernama BeOpen.com dan membentuk BeOpen PythonLabs. Setelah BeOpen mengeluarkan versi 2.0, Guido dan beberapa pengembang PythonLabs pindah ke DigitalCreations.

Sampai saat ini Guido tetap mengkoordinir pengembangan Python bersama Python Software Foundation. Python Software Foundation adalah sebuah organisasi nirlaba yang dibentuk untuk menjadi pemegang hak cipta intelektual. Nama Python dipilih Guido karena kecintaannya terhadap acara televisi Monty Python's Flying Circus yang mana seringkali ungkapan khas dari acara tersebut muncul dalam korespondensi antar pengguna Python.

### **2.2.2 *Beautiful Soup***

Beautiful Soup adalah library yang digunakan untuk mengidentifikasi sebuah tag HTML dan kemudian melakukan parsing data dari web page. Beautiful Soup akan membantu untuk mendapatkan special context dari sebuah web page, menghapus tag html dan menyimpan informasi. Beautiful Soup dapat digunakan sesuai dengan kebutuhan. Pengguna dapat menginisialisasi secara spesifik untuk

menggunakan HTML/XML analytic engine dengan memanggil perintah Beautiful Soup (markup, "lxml"). (Chunmei, Guomei, dan Zuojie, 2015)

### **2.2.3 Web Scraping**

Web Scraping adalah teknik untuk memperoleh informasi dari website secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari web scraping adalah untuk mencari informasi yang kemudian akan dikumpulkan. Fokus dari web scraping adalah untuk memperoleh data dengan cara pengambilan dan ekstraksi. Web scraping mempunyai manfaat antara lain yaitu agar informasi yang diambil lebih terfokus sehingga mempermudah dalam melakukan pencarian sesuatu. (Ayani, DD., Helen S. Pratiwi, Hafiz M. 2019)

Web scraping memiliki beberapa langkah meliputi:

1. Create Scraping Template: Pembuat program mempelajari dokumen HTML dari website yang akan diambil informasinya terkait tag HTML yang akan diambil informasinya.
2. Explore Site Navigation: Pembuat program mempelajari teknik navigasi pada website yang akan diambil informasinya.
3. Automate Navigation and Extraction: Berdasarkan informasi yang didapat pada langkah 1 dan 2, aplikasi web scraping dibuat untuk mengambil informasi secara otomatis dari website yang ditentukan.
4. Extracted Data and Package History: Informasi yang didapat dari proses nomor 3 akan disimpan dalam tabel database.