

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Penelitian-penelitian yang pernah dilakukan dengan menggunakan teknik *scraping* salah satunya adalah oleh Suhartono (2017) di Universitas Gajah Mada. Pada penelitian yang dilakukan membuat suatu *prototype* aplikasi yang diberi nama Sipeha yang akan melakukan teknik *web scraping* pada *website e-commerce* yang sudah di tentukan. Dengan menggunakan aplikasi Sipeha ini dapat menghasilkan suatu perbandingan harga pada setiap *e-commerce* yang sudah di *scraping* dan menampilkan harga dan informasi produk secara *up to date* dan menbandingkan *response time* pada dengan sistem pembanding harga yang sudah ada. Dalam penelitina ini menggunakan bahasa pemrograman PHP dengan *framework Codeigniter*. Dengan menggunakan teknik *web scraping* menggunakan *library simple HTML DOM parser*.

Penelitian kedua dilakukan Priyanto, dkk (2018) mengenai Implementasi *Web scraping* Dan *Text mining* Untuk Akuisisi Dan Kategorisasi Informasi Laman Web Tentang Hidroponik. Pededakatan khusus yang digunakan dalam penelitian ini adalah dengan menggunakan teknik *scraping* untuk akuisisi data dan *text mining* untuk kategorisasi data artikel hidroponik yang didapatkan dari laman wen yang bias diakses secara bebas. Langkah awal yang harus dilakukan adalah dengan mengakuisisi data dengan menggunakan teknik *web scraping* dari laman web yang sudah ditentukan, kemudian langkah selanjutnya adalah ekstraksi

informasi dari laman sumber menggunakan teknik *web scraping*. Penelitian ini menggunakan bahasa pemrograman *python* untuk mempermudah proses *scraping*. Kemudian informasi maupun artikel hidroponik yang berhasil diakuisisi selanjutnya dikelompokkan dengan menggunakan teknik *text mining*. Dalam *text mining* terbagi dalam dua macam yaitu *text pre-processing* dan *topic modeling*.

Penelitian ketiga yang dilakukan oleh Wijaya R.C, dkk (2015) mengenai Pembuatan Aplikasi Untuk Melakukan Perbandingan Sebuah Produk Yang Didapat Dari Beberapa Situs Web *E-commerce*. Dalam penelitian ini tidak hanya harga saja yang dibandingkan tetapi juga mengambail detail, nilai rating nam penjual juga link situs web mengenai produk yang dicari oleh pengguna. Aplikasi yang digunakan dalam penelitian ini menggunakan bahasa pemrograman PHP dan pengumpulan data dengan metode *web scraping* yang dikombinasikan dengan *library PHP simple HTML. DOM* untuk melakukan ekstrasi dari situs web *e-commerce* dan untuk *database* yang digunakan adalah *MySQL*.

Penelitian keempat dilakukan oleh Beny Fajar Riyanto (2018), penelitian ini membahas tentang *Web scraping* dan Pembobotan Kata TF-IDF untuk mencari dan mengekstrak informasi lowongan pekerjaan. Dengan menggunakan metode TF-IDF tersebut berfungsi untuk melakukan pencarian dengan mencari kemiripan data dengan kata kunci yang dimasukan, sehingga data yang diperoleh lebih baik karena dilakukan pencarian kemiripan perkata antara data lowogan dengan kata kunci yang dimasukan.

Penelitian kelima dilakukan oleh Bayu Rima Aditya (2015) di Universitas Telkom yang membahas tentang Penggunaan *Web crawler* Untuk Menghimpun *Tweets* dengan Metode *Pre-Processing Text mining* untuk mendapatkan data *tweets* pada media sosial *twitter*. Penelitian ini menguraikan hasil penggunaan *engine web crawl* menggunakan metode *pre-processing text mining*. Penggunaan *engine web crawl* bertujuan untuk menghimpun *tweets* melalui *API twitter* sebagai data teks tidak terstruktur yang kemudian direpresentasikan kembali kedalam bentuk web. Sedangkan penggunaan metode *pre-processing* bertujuan untuk menyaring *tweets* melalui tiga tahap, yaitu *cleansing*, *case folding*, dan *parsing*. Hasil dari penelitian ini adalah berupa aplikasi yang dapat mengubah 11 *tweets* yang telah dihimpun menjadi data yang siap diolah lebih lanjut sesuai dengan kebutuhan *user* berdasarkan kata kunci dan tanggal pencarian.

Dari ke lima referensi tersebut, ditemukan adanya perbedaan penerapan *web scraping* yang sudah dibuat dengan yang akan dibuat yaitu implementasi yang dikhususkan untuk menyaring informasi atau data dari *website distributor* maupun *marketplace* yang nantinya akan disajikan didalam sebuah *website* yang dapat memudahkan pengguna dalam memilih perangkat jaringan sesuai dengan harga yang diinginkan dengan menggunakan bahasa pemrograman *Python* dan *database MySQL*.

Perbandingan dengan penelitian sebelumnya, yang terkait dengan penelitian yang akan dilakukan dapat dilihat pada tabel 2.1 :

Tabel 2.1 Tinjauan Pustaka

| No | Peneliti | Objek | Tujuan penelitian | Teknik yang digunakan | Informasi yang dihasilkan |
|----|---------------------------|---|---|--|--|
| 1 | Suhartono (2017) | Data dan informasi produk dari <i>website e-commerce</i> meliputi tokopedia, bukalapak dan lazada | Membandingkan harga dan informasi produk secara <i>up to date</i> | Teknik <i>scraping</i> | Hasil perbandingan harga yang <i>up to date</i> dan respon sistem yang lebih cepat dengan <i>e-commerce</i> aslinya |
| 2 | Priyanto, dkk. (2018) | <i>Website</i> yang menyediakan artikel tentang hidroponik. http://belajarberkebun.com , http://hidroponikshop.com , dll. | Mengekplorasi laman-laman yang menyajikan informasi mengenai tatacara bercocok tanam dengan teknik mengumpulkan semua informasi yang ada di halaman tersebut secara otomatis dengan menggunakan teknologi <i>web scraping</i> | <i>Web scraping, text meaning, text analytic</i> | Kumpulan artikel mengenai <i>hidroponik</i> yang tersebar di berbagai laman web dapat diakuisisi secara otomatis kemudian dikelompokkan menjadi 12 topik menggunakan algoritma topic modelling yakni <i>Latent Dirichelet Allocation (LDA)</i> |
| 3 | Wijaya R.C (2015) | Data <i>mobile device</i> dari <i>website</i> yang menjual <i>mobile device</i> seperti www.blibli.com , www.erafone.com , www.lazada.com , dll. | Membuat aplikasi yang digunakan untuk pencarian data dan membandingkan data produk dari berbagai situs <i>e-commerce</i> dengan teknik <i>scraping</i> | Teknik <i>web scraping</i> dengan <i>library</i> PHP yaitu <i>DOMXPath</i> . | Sebuah aplikasi pencarian produk yang digunakan untuk mencari produk <i>mobile device</i> |
| 4 | Beny Fajar Riyanto (2018) | Lowongankerja pada jobs.id dan topkarir | Membangun sebuah sistem pencarian lowongan kerja, yang dapat mencari beberapa lowongan kerja dalam satu proses pencarian. | <i>Web scraping</i> dan Pembobotan kata TF-IDF | Sistem dapat melakukan pencarian dan <i>extract</i> data beberapa lowongan dalam satu proses cari. |

| | | | | | |
|---|-------------------------|---|---|-------------------------------|---|
| 5 | Bayu Rima Aditya (2015) | <i>Tweets</i> pada media sosial <i>twitter</i> | Menghimpun <i>tweets</i> melalui API <i>twitter</i> sebagai data teks tidak terstruktur yang kemudian direpresentasikan kembali kedalam bentuk web. | Penggunaan <i>Web crawler</i> | Aplikasi yang dapat mengubah <i>tweets</i> yang telah dihimpun menjadi data yang siap diolah lebih lanjut sesuai dengan kebutuhan <i>user</i> berdasarkan kata kunci dan tanggal pencarian. |
| 6 | Yang disulkan | Scraping website www.citraweb.com , www.spectrumindo.com dan www.tokopedia.com | Mengimplementasikan dasar-dasar <i>web scraping</i> | Teknik <i>Scraping</i> | Suatu <i>Website</i> yang menampilkan daftar harga perangkat jaringan dari berbagai <i>website distributor</i> maupun <i>marketplace</i> dengan teknik <i>scraping</i> |

2.2. Dasar Teori

2.2.1. *Web scraping*

Web scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman *web* dalam bahasa markup seperti *HTML* atau *XHTML* dan menganalisis dokumen tersebut untuk untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain (Turland, 2010)..

Tujuan dari *web scraper* adalah untuk mencari informasi tertentu, mengekstraksi informasi tersebut, kemudian menghimpunnya dalam halaman *web* yang baru. Beberapa *web scraper* memfokuskan dirinya untuk mengubah data yang tidak terstruktur dan menyimpannya ke dalam data yang terstruktur.

2.2.2. Perangkat jaringan

Pengertian Perangkat jaringan komputer adalah sebuah *hardware* yang di gunakan untuk mencapai suatu tujuan dari fungsi jaringan komputer itu sendiri, seperti berkomunikasi, bertukar data dan lain sebagainya (Adi Catur Pamungkas, 2019). Jenis perangkat jaringan adalah sebagai berikut :

a. *Router*

Router berfungsi untuk menghubungkan dua jaringan atau lebih, sehingga dapat saling bertukar data dengan jaringan lainnya. Maka dari itu dengan menggunakan *router* kita dapat terhubung dengan jaringan yang berbeda. Misal, IP 192.168.10.0/24 dapat terhubung dengan IP 100.100.100.0/24. Cara kerja *router* yaitu meneruskan paket data dan membagi beberapa jaringan menjadi beberapa segmen ataupun menyatukan segmen – segmen jaringan tersebut, tetapi *router* hanya terletak di lapisan ketiga pada *layer OSI*.

b. *HUB*

Hub bertugas untuk mengubah sinyal *transmisi* jaringan, hal tersebut dilakukan agar dua komputer atau lebih dapat saling terhubung. Tetapi *Hub* tidak dapat mengatur jalanya suatu data, sehingga setiap paket data yang melewati *hub* akan di *broadcast* ke semua *port* sampai paket data yang melewati *hub* akan sampai ke tujuan. Hal tersebut dapat menyebabkan adanya *collision* atau tabrakan arus data.

c. *Switch*

Switch memiliki fungsi yang hampir sama dengan *hub*, tetapi perangkat jaringan ini lebih cerdas dari *hub* karena dapat mengatasi masalah *collision* data. Selain itu *switch* juga memiliki beberapa kelebihan di antaranya yaitu memiliki kecepatan *transfer* data yang cepat dan memiliki jangkauan yang lebih baik daripada *hub*. Tidak hanya itu, *switch* juga dapat menyaring atau memfilter paket data sebelum di teruskan ke jaringan yang di tuju.

d. *Access Point*

Perangkat jaringan *Access Point* terdiri dari dua bagian utama yaitu antenna dan *transceiver* yang di gunakan untuk menerima sinyal dan *transmisi* dari *client* atau sebaliknya. Dengan *access point*, kita dapat terhubung dalam jaringan LAN searah *wireless* (tanpa kabel). Intinya, *access point* berfungsi untuk menghubungkan dua jenis jaringan yang berbeda yaitu jaringan *wireless* dan jaringan LAN.

e. *Bridge*

Bridge juga memiliki fungsi yang dibidang mirip dengan *router*, yaitu memperluas sebuah jaringan dan sekaligus membuat sebuah segmen jaringan. Cara kerja *bridge* yaitu mengenali sebuah alamat *MAC* yang akan mentransmisi sebuah

data ke jaringan, selanjutnya *bridge* akan memuat kabel *internal* secara otomatis yang dapat menentukan segmen mana yang akan di filter ataupun di *routing*.

2.2.3. Marketplace

Marketplace adalah model bisnis yang mana website yang bersangkutan tidak hanya membantu mempromosikan barang dagangan saja, tapi juga memfasilitasi transaksi uang secara *online*. Pada sistem belanja *online*, suatu website menyediakan lahan atau tempat bagi para penjual untuk menjual produk-produknya.

Tokopedia merupakan salah satu marketplace berbasis digital terbesar di Indonesia. Sejak resmi diluncurkan, PT. Tokopedia berhasil menjadi salah satu perusahaan internet Indonesia dengan pertumbuhan yang pesat. Dengan mengusung model bisnis marketplace dan mall *online*, Tokopedia memungkinkan setiap individu, toko kecil dan brand untuk membuka dan mengelola toko *online*. Sejak diluncurkan sampai hingga akhir 2015, layanan dasar Tokopedia yang bisa digunakan oleh semua orang secara gratis.

2.2.4. Tools

Adapun beberapa *tools* yang digunakan untuk membuat aplikasi *web scraping* diantaranya adalah sebagai berikut :

a. HTML

HTML adalah singkatan dari *Hypertext Markup Language*. *HTML* memungkinkan seorang user untuk membuat dan menyusun bagian paragraf, *heading*, link atau tautan, dan *blockquote* untuk halaman web dan aplikasi. *HTML* bukanlah bahasa pemrograman, dan itu berarti *HTML* tidak punya kemampuan untuk membuat fungsionalitas yang dinamis. Sebagai gantinya, *HTML* memungkinkan user untuk mengorganisir dan memformat dokumen, sama seperti *Microsoft Word* (Ariata.C, 2018).

b. PostgreSQL

PostgreSQL adalah perangkat lunak *opensource* untuk *database* atau lebih tepatnya *Object-Relational Database Management System (ORDBMS)*, *PostgreSQL* mendukung sistem operasi Linux, FreeBSD, Windows dan juga Mac OS. *PostgreSQL* bersifat *opensource* sehingga pemintanya cukup banyak dan *PostgreSQL* juga cukup stabil, handal, dan memiliki akurasi yang tinggi untuk penyimpanan basis data. (Anton Subagia, 2019)

c. Python

Python adalah salah satu bahasa pemrograman tingkat tinggi yang bersifat *interpreter*, *interaktif*, *object-oriented* dan dapat beroperasi di hampir semua *platform*, seperti keluarga *UNIX*, *Mac*, *Windows*, dan lainnya. sebagai bahasa tingkat tinggi, *Python* termasuk salah satu bahasa pemrograman yang mudah untuk dipelajari karena *sintaks* yang jelas dan elegan, dikombinasikan dengan penggunaan module-module siap pakai dan struktur data tingkat tinggi yang efisien (Dewi Rosmala, 2012).

d. Object Relational Mapping (ORM)

Object relational mapping adalah sebuah teknologi yang menjadi satu lapis antara aplikasi dengan *database*. ORM bekerja seperti *database*, tetapi hanya berbentuk objek. Setiap objek di dalam ORM umumnya mewakili tabel dalam *database*. ORM menciptakan sebuah objek basisdata yang bersifat virtual yang dapat dikenali dan digunakan dalam bahasa pemrograman. Keuntungan dari penggunaan sebuah *tool* ORM adalah untuk menghemat waktu, menyederhanakan pengembangan aplikasi dan skabilitas. (Theadoron Mayaut, 2010)