**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Perkembangan teknologi informasi yang sangat pesat saat ini, mengakibatkan perkembangan pengguna internet di Indonesia tumbuh semakin meningkat. Pada tahun 2017 pengguna internet di Indonesia meningkat 8% menjadi 143, 26 juta pengguna. Jika dibandingkan dengan data populasi penduduk Indonesia yang sebanyak 262 juta jiwa, maka pengguna internet di Indonesia sudah mencapai 54,68% dari total populasi di Indonesia. Perkembangan ini membawa pengaruh terhadap berbagai aspek peradaban, mulai dari sosial, pendidikan, kesehatan, administrasi, dan masih banyak lagi (Tim APJII, 2018).

Dari sekian banyak aspek, ekonomi merupakan aspek yang sangat dipengaruhi oleh peningkatan penggunaan internet. Dengan banyaknya penduduk Indonesia yang menggunakan internet, membuat model pertumbuhan ekonomi yang semula masih konvensional kini semakin mengarah kepada perekonomian yang berbasis internet. Salah satu bentuk pemanfaatan internet terhadap aktifitas perekonomian adalah toko *online (online shop).* *Online shop* atau yang familiar disebut *olshop* hadir ditengah masyarakat didasarkan dari kecendrungan masyarakat dalam pemanfaatan internet pada aspek kehidupan. Bahkan saat ini hampir semua bentuk kegiatan jual beli dapat dilakukan hanya dengan menggunakan perangkat elektronik dan internet. Sehingga saat ini transaksi jual beli konvensional mulai tergeser dengan transaksi jual beli modern, maka muncullah istilah perdagangan elektronik yang biasa disebut *electronic commerce (e-commerce).*

*E-commerce* sangat menguntungkan bagi produsen dan juga sangat memudahkan bagi konsumen, karena sistem pelayanan melalui internet sangat efisien, praktis dan dapat menghemat waktu untuk memasarkan produk atau jasa. Kemudahan ini menyebabkan hadirnya berbagai macam toko yang tidak memiliki bentuk fisik yaitu toko-toko maya di dunia internet atau yang dewasa ini disebut *electronic marketplace (e-marketplace)*.

*E-marketplace* yang kian merebak di dunia maya membuat persaingan untuk mendapatkan pengunjung menjadi meningkat. Beberapa *e-marketplace* mengandalkan potongan harga atau menawarkan barang dagangan dengan harga yang murah untuk menarik minat pengunjung. Namun dengan kian banyaknya *e-marketplace* yang ada saat ini membuat pengunjung kesulitan dalam menentukan atau membandingkan suatu barang yang sama pada *e-marketplace* yang berbeda.

Agar memudahkan pencarian barang pada sebuah *e-marketplace*, biasanya terdapat sistem pencarian atau mesin pencarian pada laman *website* *e-marketplace* yang dapat memfilter data sesuai dengan kata kunci yang dimasukkan oleh pengunjung. Penggunaan mesin pencarian dianggap lebih efisien dan cepat dibandingkan dengan memberikan menu *list* kepada pengunjung untuk menentukan barang yang dituju. Namun untuk memadukan data dari berbagai *website* *e-marketplace* yang ada, dapat dilakukan dengan menggunakan teknik *Web Scraping*. *Web Scraping* sendiri adalah pengambilan sebuah dokumen semi terstruktur dari berbagai sumber, yang umumnya berupa halaman-halaman web dalam bahasa markup, yang kemudian dianalisis untuk mengambil data tertentu. Dengan adanya sistem pencarian menggunakan teknik *web scraping*, *search engine* dapat melakukan pencarian *string* *input*-an terhadap beberapa *website* sekaligus.

Dalam menyempurnakan sebuah sistem pencari yang dapat memuat data dari berbagai sumber dibutuhkan sebuah algoritma yang dapat menganalisis *string*. *Levenshtein Distance* merupakan sebuah algoritma yang dapat mengukur kesamaan antara *string*. Penggabungan antara teknik *Scraping* dan algoritma *Levenshtein Distance* ini yang nantinya akan membantu pengunjung dalam menemukan dan membandingkan barang yang diinginkan.

Melihat permasalahan diatas, maka peneliti mengembangkan sebuah sistem yang mampu melakukan pencarian barang terhadap beberapa *e-marketplace* yang ada, agar dapat melakukan perbandingan atas barang yang sama terhadap *e-marketplace* yang berbeda. Atas dasar inilah peneliti mengangkat judul ***“*Implementasi *Levenshtein Distance* pada Aplikasi Pencarian Barang di Berbagai *E-Marketplace* Menggunakan Teknik *Web Scraping”*.**

* 1. **Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana menggunakan teknik *web scraping* agar mendapatkan data barang pada Bukalapak, Tokopedia dan FJB Kaskus?
2. Bagaimana mengimplementasikan algoritma *Levenshtein Distance* beserta kompleksitasnya?
3. Bagaimana cara mengujian pada sistem pencarian barang?
   1. **Batasan Masalah**

Adapun hal-hal yang menjadi batasan masalah pada penelitian ini adalah sebagai berikut :

1. Menggunakan Algoritma *Levenshtein Distance* sebagai algoritma pencarian barang pada aplikasi pencarian di berbagai *e-marketplace*.
2. Fungsi pencarian barang tidak dari semua *e-marketplace* tetapi hanya mengumpulkan dari Bukalapak, Tokopedia dan FJB Kaskus.
3. Data dari *e-marketplace* akan disimpan kedalam *database*.
4. Dibangun berbasis *website*.
   1. **Tujuan Penelitian**

Berdasarkan rumusan masalah yang sudah disebutkan sebelumnya, maka tujuan dari penelitian ini adalah untuk mengimplementasikan *Levenshtein Distance* pada aplikasi pencarian barang di berbagai *e-marketplace* menggunakan teknik *web scraping*.

* 1. **Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini dapat memberi keuntungan bagi pihak pembeli dan penjual. Dengan pencarian ini, pembeli dapat dengan mudah melakukan pencarian barang di berbagai *e-marketplace* tanpa harus membuka banyak *e-marketplace* secara langsung. Untuk keuntungan penjual, tidak perlu lagi memasang produk yang sama di beberapa *e-marketplace*.

* 1. **Sistematika Penulisan**

Sistematika penulisan tugas akhir digunakan untuk menjelaskan penulisan per bab sebagai berikut:

**BAB I PENDAHULUAN**

Merupakan bab pendahuluan yang menguraikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

**BAB II LANDASAN TEORI**

Membahas mengenai dasar-dasar teori pendukung yang berhubungan dengan masalah yang diambil dan program aplikasi yang akan digunakan dalam pembangunan sistem.

**BAB III METODE PENELITIAN**

Membahas mengenai waktu dan tempat penelitian, metode pengumpulan data, metode pembangunan sistem yang digunakan dalam penelitian tugas akhir ini.

**BAB IV ANALISIS DAN PERANCANGAN**

Bab ini akan membahas tentang analisis dan perancangan dari aplikasi yang akan dibuat, dengan menggunakan desain UML (*Unified Modelling Language*) serta desain *user interface*.

**BAB V IMPLEMENTASI DAN PEMBAHASAN**

Dalam bab ini akan dikaji mengenai implementasi hasil perancangan aplikasi yang dibuat serta melakukan pengujian terhadap sistem.

**BAB VI PENUTUP**

Bab ini berisi kesimpulan dari program yang telah dibuat serta saran yang diperlukan untuk pengembangan program berikutnya.

* 1. **Tinjauan Pustaka**

Penelitian yang dilakukan oleh Ariyani, dkk (2016) menyimpulkan bahwa pengujian menggunakan data *real* yaitu data dokumen berplagiat yang diambil dari artikel/berita lewat internet, Algoritma *Levenshtein Distance* menghasilkan nilai *similarity* yang tinggi yaitu diatas 85 % sampai 100 % untuk dokumen yang tingkat kemiripannya tinggi. Sedangkan untuk dokumen dengan tingkat kemiripan yang rendah atau tidak berplagiat maka menghasilkan nilai *similarity* dibawah 40%.

Pratama dan Pamungkas (2016) menyimpulkan bahwa Algoritma *Levenshtein Distance* untuk menghitung jumlah perbedaan antara dua *string*. Perhitungan jarak antara dua *string* ini ditentukan dari jumlah minimum operasi perubahan untuk mendeteksi kemiripan suatu dokumen teks dengan dokumen teks lainnya dengan cara pembentukan suatu matriks *string* untuk mendapatkan nilai *distance*, kemudian melakukan perhitungan bobot *similarity* antar dokumen teks berdasarkan nilai *distance* tersebut. Pada proses *sorting* juga menjadi sangat berpengaruh bagi algoritma *Levenshtein distance* dalam mendeteksi kemiripan dokumen teks. Dengan kata lain, Algoritma *Levenshtein distance* akan bekerja lebih baik jika kedua dokumen yang dibandingkan mempunyai urutan kata yang sama. Selain itu penggunaan proses *stopword*, *stemming*, dan *sorting* sekaligus dapat memberikan hasil yang lebih spesifik. Artinya, dalam mendeteksi kemiripan dokumen teks, akan lebih baik jika dilakukan pengeliminasian kata-kata yang umum, dan megubah kata berimbuhan menjadi kata dasar.

Panjaitan (2018) menyimpulkan bahwa dengan melakukan implementasi *Web Scraping* dapat mengumpulkan data melalui *website* dimana metode ini melakukan ekstraksi elemen *tag website* menjadi data yang dapat disimpan ke dalam database, sehingga dapat membentuk sebuah model pengumpulan data cuaca yang *termonitoring* dengan baik dan efisien. Identifikasi keadaan cuaca menjadi lebih akurat dengan memanfaatkan data yang besar dari hasil *scraping* dengan dikombinasikan dengan *K-Nearest Neighbor*. *K-Nearest Neighbor* dapat melakuan *resampling* dari data *historis* yang diaplikasikan untuk menentukan kondisi cuaca berdasarkan skenario cuaca yang memungkinkan. Sehingga menghasilkan akurasi identifikasi sebesar 98%.

Berdasarkan penelitian-penelitian yang telah dipaparkan di atas, maka akan dilakukan penelitian lebih lanjut dengan judul **“Implementasi *Levenshtein Distance* pada Aplikasi Pencarian Barang di Berbagai *E-Marketplace* Menggunakan Teknik *Web Scraping”.***