

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pembelian barang melalui toko daring saat ini sudah menjadi hal umum dan sudah dilakukan oleh banyak orang. Hal ini dapat dilihat berdasarkan laporan berjudul *Technology-empowered Digital Trade in Asia Pacific* dari Deloitte bahwa total besar market *e-commerce* di Indonesia tercatat mencapai US\$ 43,351 miliar pada 2021, belum lagi tingkat konsumerisasi negara Indonesia yang cukup tinggi pada pasar berkembang [1]. Hal ini membuat banyak *e-commerce* baik dari luar negeri maupun lokal, mulai bersaing di pasar Indonesia, seperti Shopee, Tokopedia, JD, Blibli, dan lain-lain. Hal ini membuat budaya belanja daring mulai menjamur, dari semua kalangan, dari anak-anak hingga dewasa juga sudah mulai belanja daring untuk memenuhi kebutuhan mereka.

Meningkatnya kegiatan belanja daring ini membuat orang-orang sering belanja daring tanpa hati-hati. Pasalnya banyak juga pedagang menipu pembelinya dengan berbagai cara, seperti barang yang diantarkan tidak sesuai, barang yang datang dalam kondisi buruk, barang yang datang tidak memiliki kualitas yang baik, dan lain-lain. Maka dari itu sebagai pembeli harus sangat teliti dalam membeli, mulai dari melihat deskripsi secara detail, membandingkan satu barang dengan barang lain di lapak yang lain, membeli melalui lapak yang terpercaya, ataupun melihat ulasan barang dari orang-orang yang telah membelinya. Namun tak sedikit juga pedagang mencurangi ulasan agar barang tersebut tampak terlihat baik, sehingga pembeli tertarik untuk membeli barang tersebut tanpa berpikir panjang. Maka dari itu saya sebagai penulis tertarik untuk membangun sebuah sistem yang dapat mempelajari pola ulasan yang baik dan benar kemudian akan

diaplikasikan untuk mendeteksi ulasan palsu agar dapat menjadi referensi pengguna atau pembeli ketika ingin membeli suatu barang.

Teknologi yang dipakai untuk membuat sistem yang dapat mendeteksi ulasan palsu adalah teknologi pemroses bahasa natural atau *natural language processing* (NLP). *Natural language processing* atau NLP merupakan area yang membahas bagaimana komputer dipakai untuk memahami dan memanipulasi bahasa natural (bahasa manusia) baik itu secara teks maupun bicara yang nantinya dapat dipakai untuk hal yang lebih bermanfaat lagi [2]. Dalam kasus ini akan diambil 2 data dari 2 *e-commerce* berbeda berupa ulasan untuk setiap produk, lalu ulasan akan dinilai terindikasi palsu atau terindikasi asli (original). Penilaian ini kemudian akan dijadikan referensi untuk sistem dan menentukan apakah ulasan tersebut palsu atau asli. Palsu disini bisa dibilang sebagai terindikasi bahwa ulasan tersebut dibuat-buat dan tidak di dapat dari pelanggan yang telah membelinya.

Untuk dataset yang digunakan adalah dataset Tokopedia dan Shopee, hal ini dilakukan untuk mendeteksi apakah ada kesamaan ulasan pada kedua *e-commerce* tersebut. Data yang diambil dari dataset berupa ulasan, barang, pengguna serta waktu ulasan dalam bentuk *timestamp*, dll. Untuk setiap dataset yang diambil (baik itu menggunakan Shopee maupun Tokopedia) dipisah menjadi dataset yang sudah dilabeli (sebanyak 100 data) dan dataset yang dilakukan untuk pengujian. Setelah itu dataset yang sudah dilabeli akan dicari nilai kesamaan melalui *Cosine Similarity* untuk menentukan kesamaan antara data yang sudah dilabeli sehingga nantinya akan sangat membantu untuk menentukan apakah ulasan tersebut asli atau palsu.

Berdasarkan uraian diatas, dalam penelitian ini akan dibuat model untuk mendeteksi ulasan palsu yang berdasar pada 2 *e-commerce* yaitu Tokopedia dan Shopee. Adapun algoritma yang digunakan adalah *Logistic Regression* yang digunakan untuk mengklasifikasi apakah ulasan tersebut terindikasi asli atau terindikasi palsu. Oleh karena itu peneliti membuat

penelitian yang berjudul “Deteksi Ulasan Palsu pada Tokopedia dan Shopee Menggunakan *Natural Language Processing*”.

Harapannya dengan adanya penelitian ini mampu menambahkan *awareness* konsumen ketika membeli barang pada *e-commerce*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah didalam penelitian ini adalah sebagai berikut:

1. Bagaimana model yang terbentuk dari hasil pelatihan pada deteksi ulasan palsu menggunakan *natural language processing* menggunakan algoritma *logistic regression*?
2. Bagaimana membuat model yang dapat melakukan analisa ulasan lintas platform?
3. Bagaimana hasil akhir dari model yang telah terbentuk dan diterapkan ke *website*?

1.3 Batasan Masalah

Berikut batasan-batasan yang ada dalam penelitian ini dengan dimaksudkan agar penelitian lebih rapi dan terarah:

1. Penelitian ini tidak melakukan implementasi pada *dummy website* Tokopedia dan Shopee, melainkan menggunakan dataset Tokopedia dan Shopee sebagai dasar untuk pelatihan *machine learning*.
2. Bahasa pemrograman yang digunakan adalah python dengan *library* yang menundukung proses pembuatan sistem *natural language processing*.
3. Data yang digunakan hanya data *review*, *item*, serta *user* dan yang terkait.

4. Dataset diambil tidaklah real time melainkan data *scraping* pada saat itu .
5. Algoritma yang digunakan adalah *logistic regression*.
6. Penelitian tidak mencakup analisa akun pengguna *real time* yang memberikan ulasan terhadap suatu produk, sehingga analisisnya hanya sebatas pengguna yang tercantum pada dataset saja.
7. Label palsu atau asli memiliki makna terindikasi, artinya ulasan tersebut kemungkinan tinggi asli atau palsu.
8. Pada penelitian ini tidak juga menggunakan deteksi bot sebagai verifikasi.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui model yang terbentuk dari hasil pelatihan pada deteksi ulasan palsu menggunakan algoritma *Logistic Regression*.
2. Membuat model yang dapat melakukan generalisasi sehingga model tersebut dapat menganalisa ulasan lintas platform.
3. Membangun sistem berbasis GUI yang dapat mendeteksi ulasan palsu secara otomatis

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Dapat memudahkan konsumen untuk melakukan *cross check* terhadap ulasan palsu yang ada di lapak dua *e-commerce*.
2. Dengan diketahuinya jenis ulasan yang palsu yang berpotensi untuk menjatuhkan lapak pedagang, sehingga dapat memudahkan pedagang untuk memilah mana kritik yang sifatnya membangun yang nantinya akan jadi acuan pedagang untuk meningkatkan kualitas produknya.

3. Hasil penelitian ini dapat dijadikan acuan lebih lanjut yang berbasis pada *natural language processing* menggunakan *logistic regression*.

1.6 Sistematika Penulisan

Sistematika penulisan tugas akhir ini disusun sebagai berikut:

Bab I Pendahuluan, berisi: latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, serta sistematika penulisan.

Bab II Landasan Teori, berisi: hasil penelitian sejenis yang sudah pernah dilakukan sebelumnya, teori penunjang, dan referensi berupa buku, jurnal, dan laporan skripsi/tesis.

Bab III Metodologi Penelitian, berisi: penjelasan mengenai metode penelitian yang digunakan untuk memahami dan mengeksplorasi obyek penelitian, hasil observasi / pengumpulan data, masalah yang terdapat pada obyek, dan gambaran umum proyek atau obyek penelitian, hingga Rencana Alur Penelitian.

Bab IV Pembahasan, berisi: rancangan proyek, implementasi coding dan desain, serta evaluasi rancangan. Selanjutnya alur pengerjaan proyek, metode testing, hingga hasil akhir penelitian dan pembahasan analisis hasil akhir penelitian, termasuk pembahasan hasil-hasil uji coba (testing). Data hasil akhir pengujian dapat berupa grafik, table, data monitoring, log system, dan lain-lain, dengan pembahasan.

Bab V Penutup, berisi kesimpulan dari hasil akhir penilaian proyek, dan saran.