

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Banyaknya pengguna internet dan media sosial di Indonesia menyimpan potensi besar yang bisa dimanfaatkan untuk melakukan transaksi penjualan secara *online*. Hal tersebut sejalan dengan meningkatnya minat berbelanja masyarakat Indonesia, yang didukung oleh perubahan perilaku konsumen untuk melakukan transaksi digital. GlobalWebIndex [1] melaporkan bahwa Indonesia memiliki tingkat tertinggi penggunaan *e-commerce* di antara negara manapun di dunia, dengan 90 persen pengguna internet di negara tersebut yang berusia antara 16 dan 64 tahun melaporkan bahwa mereka telah membeli produk dan layanan secara online. Aktivitas *e-commerce* di Indonesia dapat dilihat pada Gambar 1.1.



Gambar 1.1 *E-commerce Activities In Indonesia*

Hadirnya *e-commerce* memungkinkan masyarakat untuk berbelanja dimana saja dan kapan saja melalui media internet. Salah satu model bisnis *e-commerce* yaitu *e-marketplace*. *E-marketplace* merupakan sebuah website atau

aplikasi jual beli online yang memiliki banyak *vendor*, model bisnis yang digunakan adalah *Customer to Customer* (C2C). Gambar 1.2 merupakan data transaksi *e-commerce* di Indonesia berdasarkan kategori. Menurut data awal tahun 2019 tersebut, pembelanjaan *online* konsumen di Indonesia paling banyak pada kategori *travel*, perangkat elektronik, dan *fashion* [1].



**Gambar 1.2 E-commerce Spend By Category**

Dengan semakin bertambahnya jumlah penjual dan pembeli pada *e-marketplace* menciptakan persaingan yang ketat antar penjual, sehingga penjual harus mencari cara agar produknya menarik dan dibeli. Salah satunya, dengan cara membuat judul produk yang unik, seperti menambahkan kata gratis, garansi, dan lain sebagainya. Selain itu, penjual melakukan penambahan merek terkenal lain di belakang judul produk untuk meningkatkan peringkat pada proses pencarian. Di sisi lain, calon pembeli mengalami kesulitan untuk mencari produk yang sesuai dengan keinginannya karena banyaknya produk yang dijual [2].

*Text mining*, disebut juga sebagai *Teks Data Mining* (TDM) atau *Knowledge Discovery in Text* (KDT), secara umum merujuk pada proses

ekstraksi informasi yang berguna dari dokumen-dokumen teks tak terstruktur (*unstructured*) [3]. *Text mining* dapat digunakan untuk mengklasifikasikan dokumen ke dalam suatu kategori. Tujuan dari kategorisasi teks adalah menguji pengklasifikasian dokumen yang belum diketahui kategorinya, sehingga apabila terdapat dokumen baru akan lebih mudah diklasifikasikan pada suatu kategori berdasarkan dokumen-dokumen yang ada [4]. Salah satu pemanfaatan *text mining* pada *e-marketplace* yaitu mengelompokkan data produk yang sama agar pembeli lebih mudah ketika melakukan pencarian produk yang diinginkan.

Penelitian yang terkait dengan latar belakang permasalahan diatas sudah pernah dilakukan oleh Danny Sebastian (2019) dengan judul Implementasi Algoritma *K-Nearest Neighbor* untuk Melakukan Klasifikasi Produk dari beberapa *E-marketplace* [2]. Pada penelitian tersebut metode *K-Nearest Neighbor* dapat melakukan klasifikasi produk dari *e-marketplace*, khususnya tokopedia dan bukalapak. Berdasarkan pengujian 1, disimpulkan bahwa nilai  $k$  yang optimal adalah 5. Kemudian akurasi yang dihasilkan pada pengujian 2 adalah 96,67%. Nilai akurasi ini dapat dikatakan baik karena melebihi 90%.

Berdasarkan penelitian sebelumnya, maka dalam penelitian ini penulis melanjutkan saran pengembangan dengan melakukan pengujian ke data produk yang berasal dari *e-marketplace* selain tokopedia dan bukalapak. Klasifikasi yang dilakukan pada penelitian ini menggunakan metode *K-Nearest Neighbor*, sehingga klasifikasi dilakukan berdasarkan data *training* dilihat dari jarak yang paling dekat dengan objek berdasarkan nilai  $k$ .

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan adalah “Bagaimana cara mengimplementasikan algoritme *K-Nearest Neighbor* untuk melakukan klasifikasi produk pada *e-marketplace* Shopee?”.

## 1.3 Batasan Masalah

Agar penelitian ini tidak keluar dari pokok permasalahan yang telah dirumuskan, maka peneliti memberikan batasan masalah sebagai berikut:

1. Klasifikasi dilakukan dengan menggunakan algoritme *K-Nearest Neighbor*.
2. Data-data produk yang digunakan dalam penelitian ini diambil dari *e-marketplace* Shopee dengan cara *web-scraping*.
3. Data produk yang digunakan sebanyak 45 data yang terdiri dari 15 data produk dari kategori *handphone* dan aksesoris, 15 data produk dari kategori *fashion muslim*, dan 15 data produk dari kategori perlengkapan rumah .
4. Kelas klasifikasi pada penelitian ini terdiri dari *iphone 7 plus*, *iphone X*, *gamis*, *dress*, *rak gantung*, dan *hanger*.
5. Pembuatan program menggunakan bahasa pemrograman PHP dan *database MySQL*.
6. Penelitian ini tidak membahas kecepatan algoritme dalam klasifikasi.

#### 1.4 Maksud dan Tujuan Penelitian

Berdasarkan permasalahan di atas, tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengklasifikasikan produk pada *e-marketplace* Shopee menggunakan algoritme *K-Nearest Neighbor*.
2. Menemukan tingkat akurasi algoritme *K-Nearest Neighbor* dalam mengklasifikasikan produk pada *e-marketplace* Shopee.

#### 1.5 Manfaat Penelitian

Berdasarkan permasalahan di atas, manfaat yang diharapkan dari penelitian ini adalah:

1. Manfaat bagi peneliti

Penelitian ini bermanfaat sebagai sarana untuk menambah wawasan dan memperoleh pengetahuan mengenai pengolahan data menggunakan *text mining*.

2. Manfaat bagi pengguna

Pengguna lebih mudah dalam melakukan pencarian produk yang diinginkan.

3. Manfaat bagi objek penelitian

Perusahaan dapat meningkatkan *costumer engagement* sehingga ketika perusahaan memiliki program-program yang baru, konsumen akan lebih tergerak untuk berpartisipasi atau terciptanya *customer loyalty* dan pembelian berulang.

## **1.6 Metode Penelitian**

Dalam melakukan penelitian ini diperlukan beberapa metode agar mencapai hasil yang diharapkan yaitu sebagai berikut.

### **1.6.1 Metode Pengumpulan Data**

Dalam memperoleh data dan informasi sebagai penunjang penelitian ini, dilakukan pengambilan data dari suatu *website* secara spesifik. Dalam penelitian ini, *dataset* yang digunakan merupakan hasil *web-scraping* dari *website* Shopee.

### **1.6.2 Metode Studi Pustaka**

Metode pengumpulan data dan informasi yang diperoleh dari buku-buku, jurnal ilmiah, tesis, internet, literatur, serta sumber-sumber lain yang berhubungan dengan objek penelitian. Peneliti melakukan studi pustaka mengenai metode dan algoritme yang akan digunakan, yaitu metode TF-IDF, *cosine similarity*, *confusion matrix*, dan algoritme *K-Nearest Neighbor*.

### **1.6.3 Metode Analisis**

Pada tahap ini dilakukan analisis terhadap studi literatur untuk mendapatkan pemahaman mengenai metode yang digunakan, yaitu algoritme *K-Nearest Neighbor* dan *text mining* untuk menyelesaikan permasalahan dalam klasifikasi produk.

### **1.6.4 Metode Perancangan**

Pada tahap ini dilakukan perancangan bagaimana alur sistem ini akan berjalan. Selain itu, akan dilakukan perancangan *input* dan *output* sistem.

### **1.6.5 Metode Implementasi**

Dalam metode implementasi ini terdapat beberapa tahapan yang akan dilakukan yaitu:

1. Data hasil *scraping* akan digunakan sebagai data *training* dan data *testing*.
2. *Text preprocessing*

Tahapan dari sistem klasifikasi dimulai dengan *preprocessing*, data produk yang telah dikumpulkan akan melewati empat tahap yaitu *case folding*, *tokenizing*, *stemming*, dan *filtering*.

3. Pembobotan TF-IDF

Term-term yang terdapat pada setiap dokumen hasil *preprocessing* diberi nilai atau bobot menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), dimana nilai tersebut akan dijadikan *input* pada klasifikasi.

4. Klasifikasi algoritme *K-Nearest Neighbor*

Setelah diperoleh bobot setiap term, kemudian dihitung jarak kemiripan antar dokumen menggunakan algoritme *cosine similarity*. Selanjutnya, data *training* diurutkan berdasarkan hasil *cosine similarity* untuk diambil sejumlah nilai  $k$  data yang memiliki kedekatan.

5. Pengujian sistem

Tingkat akurasi algoritme *K-Nearest Neighbor* pada sistem ini akan dihitung menggunakan rumus *confusion matrix*.

#### 1.6.6 Metode Pengujian

Pada tahap ini dilakukan pengujian terhadap sistem untuk mengetahui seberapa baik sistem dalam melakukan klasifikasi, serta mengukur tingkat akurasi, nilai *precision*, dan nilai *recall* dari algoritme *K-Nearest Neighbor* didasarkan pada perhitungan *confusion matrix*.

## 1.7 Sistematika Penulisan

Untuk memudahkan pembaca dalam mengikuti apa yang dipaparkan dalam laporan skripsi ini, sistematika penulisan tersusun dari 5 (lima) bab yaitu:

### **BAB I      PENDAHULUAN**

Bab ini berisikan gambaran umum berupa latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan laporan penelitian.

### **BAB II     LANDASAN TEORI**

Bab ini menguraikan dasar-dasar teori yang mendukung proses pengklasifikasian produk menggunakan algoritme *K-Nearest Neighbor*. Selain itu, penelitian-penelitian sebelumnya yang digunakan sebagai acuan dalam penulisan laporan penelitian.

### **BAB III    ANALISIS DAN PERANCANGAN SISTEM**

Bab ini membahas mengenai analisis dan perancangan sistem menggunakan algoritme *K-Nearest Neighbor*. Pada tahap analisis dibahas mengenai analisis data dan analisis metode yang digunakan. Kemudian pada tahap perancangan dibahas mengenai perancangan alur sistem dan tampilan antarmuka sistem.

### **BAB IV    IMPLEMENTASI DAN PEMBAHASAN**

Bab ini berisi uraian implementasi algoritme *K-Nearest Neighbor* untuk klasifikasi produk dan pengujian terhadap sistem yang telah berhasil dibuat.

**BAB V      PENUTUP**

Bab ini merupakan bab terakhir yang berisi kesimpulan dari penelitian yang telah dilakukan. Selain itu, bab ini juga berisi saran yang berguna untuk perkembangan penelitian yang akan datang.

