BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Dalam studi terbaru yang dilakukan oleh Nasution et al. [1], dipaparkan bahwa perkembangan E-Commerce yang sangat pesat menciptakan pasar sendiri dan berdampak pada sektor bisnis digital, namun juga menimbulkan berbagai macam tantangan. Hasil studi menemukan bahwa masyarakat kesulitan memilih produk E-Commerce yang tepat karena banyaknya produk yang tersedia. Di lain sisi, Alhamra dalam penelitiannya untuk merancang aplikasi pencarian produk Sneaker, dijelaskan bahwa masyarakat kesulitan untuk memilih produk dengan harga dan kualitas terbaik sehingga perlunya teknologi pencarian produk yang lebih baik [2].

Untuk menganalisa data teks, pertama-tama diperlukan metode yang dapat memproses data teks. Metode yang diteliti untuk masalah klasifikasi teks contohnya adalah Term Frequency-Inverse Document Frequency (TF-IDF) [3]. TF-IDF digunakan untuk mengekstraksi fitur-fitur penting dari data teks dengan menghitung frekuensi kemunculan kata-kata dalam suatu dokumen dan memperhitungkan kepentingan kata-kata tersebut dalam dokumen tersebut serta keseluruhan koleksi dokumen. Dengan menerapkan metode TF-IDF, data teks dapat diwakili oleh vektor-vektor fitur yang merepresentasikan kepentingan relatif dari kata-kata dalam deskripsi tersebut. Metode TF-IDF ini telah digunakan untuk memecahkan berbagai permasalahan terkait analisa data teks, antara lain klasifikasi teks [4], analisa sentimen masyarakat terhadap kebijakan lockdown Covid-19 [5], pencarian similaritas judul tugas akhir [6], dan sentimen masyarakat tentang ekonomi kreatif [3].

Selanjutnya, setelah teks telah berhasil diubah menjadi bentuk vektor yang dapat dimengerti algoritma, maka perlu algoritma untuk mengklasifikasikan teks tersebut seperti Support Vector Machine (SVM) [5], [7], [8]. SVM adalah algoritma pembelajaran mesin yang efektif dalam menangani masalah klasifikasi dengan membangun sebuah model yang memisahkan dua kelas dengan margin maksimum. Beberapa penelitian sebelumnya telah membuktikan bahwa SVM dapat menghasilkan klasifikasi teks dengan akurasi yang tinggi untuk mengklasifikasikan teks media sosial

twitter tentang berita penusukan Wiranto [8], teks SMS sending and posting advertisement in mass (SPAM) [7], media sosial Twitter tentang lockdown Covid-19 [5], dan ulasan produk kecantikan [9].

Dataset berupa big data bertipe teks yakni teks deskripsi produk yang telah dikategorikan menjadi empat macam, yaitu "Elektronik", "Rumah Tangga", "Buku", dan "Pakaian & Aksesori", dengan total 50425 data. Dataset ini berformat ".csv" dan terdiri dari dua kolom, dengan kolom pertama berisi nama kelas dan kolom kedua berisi produk dan deskripsi yang terkait. Dataset ini merupakan representasi dari situs e-commerce, yang diambil melalui proses scraping, dan memiliki karakteristik multivariat. Tujuan utama penggunaan dataset ini adalah untuk melakukan tugas klasifikasi dengan bantuan algoritma SVM dan TF-IDF.

Solusi yang diusulkan berupa algoritma yang dirancang dengan program Jupyter Notebook berbasis bahasa pemrograman Python 3 untuk mengklasifikasikan deskripsi produk E-Commerce. Pentingnya membantu konsumen dalam memilih produk E-Commerce yang tepat memotivasi para peneliti untuk mengangkat penelitian yang dapat membantu memecahkan masalah tersebut. Selain itu dataset yang digunakan adalah deskripsi produk dan algoritma yang digunakan adalah TF-IDF dan SVM, dan mengembangkan sebuah model klasifikasi teks yang diberi judul "KLASIFIKASI DESKRIPSI PRODUK E-COMMERCE DENGAN METODE TF-IDF DAN SVM".

1.2. Rumusan Masalah

Rumusan masalah penelitian antara lain:

1. Bagaimana mengklasifikasikan deskripsi produk secara otomatis berdasarkan kategori atau atribut yang relevan menggunakan metode TF-IDF dan SVM?
2. Bagaimana mengoptimalkan representasi vektor kata-kata menggunakan metode TF-IDF agar dapat mengidentifikasi fitur-fitur penting dalam deskripsi produk dengan lebih akurat?
3. Bagaimana menghasilkan model dengan akurasi tinggi untuk klasifikasi deskripsi produk dengan menggabungkan metode TF-IDF dan SVM?

1.3. Tujuan

Tujuan yang hendak dicapai adalah menciptakan sebuah metode yang dapat membantu komputer untuk secara otomatis mengelompokkan deskripsi-deskripsi produk berdasarkan jenis atau bagian yang penting. Metode ini menggunakan metode yang disebut TF-IDF dan SVM. Dengan menggunakan kedua metode ini, diharapkan dapat membuat komputer lebih pintar dalam memahami deskripsi-deskripsi produk dan memberikan klasifikasi yang tepat. Hal ini akan membantu dalam memilih produk yang sesuai dengan kebutuhan dan preferensi secara lebih mudah.

1.4. Manfaat

Manfaat yang diharapkan dari dilakukannya antara lain:

1. Mengembangkan metode yang dapat membantu dalam mencari produk yang diinginkan dengan lebih mudah. Dengan menggunakan komputer, pengguna hanya perlu memberikan deskripsi atau kata-kata yang relevan, dan komputer akan membantu mengelompokkan produk sesuai dengan jenis atau kategori yang relevan.
2. Mengembangkan metode klasifikasi yang lebih baik sehingga komputer akan mampu memberikan rekomendasi produk yang lebih sesuai dengan kebutuhan dan preferensi pengguna. Hal ini akan membantu pengguna menghemat waktu dan usaha dalam mencari produk yang sesuai, serta meningkatkan kepuasan dalam berbelanja.
3. Memberikan manfaat bagi perusahaan dan penjual dalam menganalisis pasar dengan lebih baik. Dengan demikian diharapkan dapat memahami preferensi dan kebutuhan pelanggan berdasarkan klasifikasi produk yang dihasilkan.

1.5. Batasan Masalah

Ruang lingkup penelitian adalah sebagai berikut:

1. Produk yang diteliti sebanyak empat macam, yaitu Elektronik, Peralatan Rumah Tangga, Buku, dan Pakaian & Aksesori.
2. Menggunakan big data, yakni data bertipe teks yang terdiri dari 50425 sampel data deskripsi produk.
3. Eksperimen yang dilakukan menggunakan bahasa pemrograman Python 3 sebagai bahasa pemrograman utama untuk mengimplementasikan metode TF-IDF dan SVM.

1.6. Keterbaruan

Penelitian sebelumnya yang telah meneliti hal serupa antara lain ditampilkan pada Tabel 1.1 berikut ini.

Tabel 1. 1. Keterbaruan

No.	Penelitian Sebelumnya	Keterbaruan
1.	Kabra dan Nagar[10] Menggunakan metode TF-IDF dan SVM untuk klasifikasi berbagai hal seperti blog dan teks. Hasil penelitian menunjukan TF-IDF dan SVM dapat diandalkan untuk mengklasifikasi Blog dan teks.	Meskipun metode yang digunakan sama dengan penelitian sebelumnya yaitu TF-IDF dan SVM, namun terdapat keterbaruan dengan fokus pada klasifikasi dataset deskripsi produk E-Commerce. Perbedaan dalam dataset ini penting karena memberikan kesempatan untuk mengeksplorasi kinerja dan adaptabilitas metode TF-IDF dan SVM dalam data yang berbeda, yang mungkin menunjukkan karakteristik dan fitur linguistik yang berbeda dibandingkan dengan blog dan teks umum.
2.	Eke et al. [11] Menggunakan metode GloVe untuk mengubah kata-kata menjadi vektor yang mewakili makna dan konteks. Hasil penelitian menunjukan bahwa metode Glove efektif dalam mengubah kata-kata menjadi vektor.	Meskipun metode yang digunakan mirip dengan metode pada penelitian sebelumnya yaitu adalah metode GloVe, namun terdapat keterbaruan dari segi metode menggunakan TF-IDF yang cukup berbeda dengan metode GloVe yaitu metode TF-IDF lebih berfokus untuk mengonversi deskripsi produk E-Commerce menjadi representasi numerik. Tujuannya adalah untuk mengetahui apakah TF-IDF dapat seefektif GloVe dalam menangkap kesamaan semantik dan konteks dalam deskripsi produk, yang dapat memberikan informasi penting dalam penelitian serupa.
3.	Ali et al. [12] Menggunakan jaringan saraf LSTM untuk memahami hubungan dan pola dalam teks. Hasil penelitian menunjukan bahwa	Berbeda dengan penelitian yang dilakukan sebelumnya yang menggunakan jaringan saraf LSTM untuk memahami hubungan dan pola dalam teks, penelitian ini menggunakan algoritma SVM untuk

	<p>algoritma LSTM dapat memahami hubungan dan pola dalam teks.</p>	<p>klasifikasi. Pemilihan algoritma SVM dengan metode TF-IDF juga menjadi keterbaruan karena penelitian sebelumnya tidak menggunakan TF-IDF.</p>
4.	<p>Pratiwi et al. [9] menggunakan algoritma SVM untuk mengklasifikasikan ulasan produk kecantikan. Hasil penelitian menunjukkan bahwa algoritma SVM dapat menghasilkan akurasi yang bagus untuk mengklasifikasikan teks ulasan produk kecantikan.</p>	<p>Meskipun penelitian sebelumnya berhasil menggunakan SVM untuk mengklasifikasikan ulasan produk kecantikan, penelitian ini menerapkan metode SVM pada dataset yang berbeda, yaitu deskripsi produk E-commerce. Keterbaruan ini dapat bermanfaat untuk menganalisis bagaimana algoritma SVM, yang telah menunjukkan keunggulan dalam klasifikasi teks ulasan produk kecantikan, dapat diterapkan pada data lainnya seperti data deskripsi produk E-Commerce.</p>
5.	<p>Arifin et al. [4] mengkategorikan artikel ilmiah dengan metode SVM dan TF-IDF. Hasil penelitian menunjukkan bahwa algoritma SVM dapat menghasilkan akurasi yang cukup baik.</p>	<p>Sejalan dengan penelitian sebelumnya yang menggunakan metode SVM dan TF-IDF. Namun, terdapat keterbaruan dengan fokus pada klasifikasi teks deskripsi produk E-Commerce. Perubahan dataset dari artikel ilmiah ke deskripsi produk menawarkan percobaan yang berguna untuk mengkaji keberagaman dan keunggulan metode SVM dan TF-IDF.</p>
6.	<p>Ristiano T [8] mengklasifikasikan data posting Twitter dengan algoritma SVM. Hasil penelitian menunjukkan bahwa algoritma SVM dapat menghasilkan akurasi yang baik.</p>	<p>Penelitian saat ini menggunakan algoritma yang sama dengan penelitian sebelumnya yaitu SVM, tetapi dengan keterbaruan pada klasifikasi dataset deskripsi produk yang berbeda. Hal tersebut menjadi eksplorasi lebih lanjut tentang bagaimana SVM, yang telah terbukti efektif dalam mengklasifikasikan posting Twitter, dapat diadaptasi dan diterapkan dalam konteks E-commerce untuk deskripsi produk.</p>