

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tren *fashion* terus mengalami perkembangan pesat di Indonesia, beragam model dan jenis pakaian terus bermunculan yang disambut hangat dan antusias oleh para pencinta *fashion* (Arsita, 2022). Gaya berpakaian dipercaya dapat mencerminkan kepribadian diri, tampil menarik dengan mengenakan *fashion* yang keren dan *up to date* dinilai sebagai salah satu cara mengekspresikan diri yang dapat diperlihatkan pada khalayak umum (Holbrook dan Hirschman, 1982). Tren *fashion* di Indonesia didominasi oleh *style* Korea dan barat, walaupun ada juga yang berciri khas Indonesia. *Brand* luar juga sudah sangat terkenal dan berpengaruh kuat dalam pasar Indonesia. Meskipun begitu, saat ini brand lokal sudah mulai tumbuh dan tak kalah bersaing dengan *brand* luar yang terkenal. Eksistensi *brand* lokal semakin didukung dengan kemudahan pemasaran produknya. Sebab, sudah banyak *e-commerce* (pasar online) yang bisa dipilih saat berbelanja (Haryanti dkk., 2020). Kehadiran pasar online memudahkan para pelaku dunia *fashion* dalam memasarkan produk dengan jangkauan yang lebih luas. Para pembeli mendapatkan kemudahan dengan banyaknya pilihan produk.

Pasar *online* menjadi alternatif konsumen karena kenyamanan dan kemudahan akses dalam berbelanja. Faktor yang memengaruhi keputusan belanja pada pasar *online* salah satunya adalah ketersediaan informasi (Li dkk, 2019). Selain deksripsi produk, konsumen akan melihat ulasan konsumen lain pada produk tersebut sebagai informasi dalam memutuskan pembelian produk (Wang dkk, 2022). Setiap kalimat yang dituliskan konsumen memiliki makna dan emosi tersirat tentang *brand* yang ditawarkan, melalui ulasan ini konsumen berbagi pengalaman dan memberikan informasi tentang produk kepada pembeli baru. Jumlah data berdasarkan hasil *scrapping* ulasan yang dilakukan pada proses penelitian ini pada bulan Juli 2022, jumlah data yang terkumpul sebanyak 11.615 data kotor (<https://shopee.co.id/>).

Masalah yang muncul adalah sulitnya mengkategorikan dan mengidentifikasi makna ulasan konsumen karena banyak ulasan yang timbul dan bervariasi serta

pemberian rating bintang oleh konsumen yang terkadang dilakukan secara acak. Oleh karena itu untuk mengetahui makna dari kecenderungan informasi dalam ulasan konsumen yang berjumlah banyak, dibutuhkan suatu metode untuk mengolah data ulasan konsumen secara cepat dan otomatis. Analisis sentimen adalah metode yang dapat digunakan untuk mengetahui makna dari kecenderungan informasi.

Ulasan konsumen adalah opini dari konsumen yang sudah membeli terkait pengalaman terhadap produk yang dibeli. Ulasan konsumen sebagai sumber informasi yang dipercaya tinggi bagi calon konsumen sebelum membeli sebuah produk (Shankar dkk, 2021). Calon konsumen terbiasa membaca ulasan untuk mendapatkan informasi sebuah produk dan mengevaluasi alternatif produk lainnya. Fitur kolom komentar pada pasar *online* disediakan bagi konsumen untuk menuliskan ulasan dan memberikan nilai dengan skala pemberian bintang 1 sampai 5 terhadap suatu produk. Ulasan yang berupa deskripsi teks juga dapat digambarkan dalam bentuk skala pemberian bintang 1 sampai 5 untuk memudahkan konsumen menerjemahkan ulasan. Selain itu perlu diperhatikan, mengandalkan rating sebagai acuan untuk menentukan sentimen konsumen terhadap suatu produk tidaklah cukup karena seringkali terdapat bias di antara keduanya (Singla dkk, 2017). Ulasan yang timbul dalam bentuk teks ulasan konsumen menjadi sekumpulan teks yang membentuk sebuah data informasi ulasan konsumen yang tidak terstruktur. Data teks ulasan konsumen yang berlimpah dapat dimanfaatkan untuk menggali informasi yang lebih mendalam terkait opini pelanggan. Informasi yang membentuk data tidak terstruktur dapat diubah dengan bantuan sentimen analisis untuk menggali makna dibalik ulasan pelanggan.

Terkait penelitian terdahulu mengenai klasifikasi sentimen analisis dan metode penyeimbangan data, terdapat beberapa penelitian yang dilakukan dengan menggunakan metode *machine learning*. Hadwan dkk (2022) dalam risetnya melakukan klasifikasi sentimen analisis pengukuran kepuasan pelanggan terhadap aplikasi seluler menggunakan algoritma *Random Forest* (RF), *Bagging*, *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Naïve Bayes* (NB) dan teknik SMOTE. Hasil komparasi dari masing-masing algoritma menunjukkan skor akurasi

tertinggi diperoleh algoritma *Support Vector Machine* (SVM) dengan nilai skor akurasi 94,38% (Hadwan dkk, 2022). Arivoli dan Sonali (2021) melakukan penelitian sentimen analisis dengan dataset ulasan film Polarity yang diambil dari API *Twitter*. Algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine* digunakan dalam penelitian ini. Hasil evaluasi menggunakan *confusion matrix* dari kedua algoritma mendapatkan skor akurasi dari algoritma *Naïve Bayes* 76,67 dan skor akurasi algoritma *Support Vector Machine* 78,18 (Arivoli dan Sonali 2021). Rai dkk (2022) melakukan penelitian terkait prediksi kelainan jantung dengan metode penyeimbangan data. Algoritma *CNN-LSTM* diterapkan sebagai pembelajaran dalam proses prediksi. Metode *SMOTE-Tomek Links* digunakan untuk mengatasi masalah ketidakseimbangan data di bidang medis dan perawatan kesehatan, yang membuat performa prediksi tidak memberikan hasil yang diinginkan di semua aspek. Data EKG yang berjumlah 123.998 dari dua kumpulan data EKG berdasarkan “MITDB dan PTDBD” yang terintegrasi digunakan untuk pelatihan dan validasi model. Hasil prediksi dengan *SMOTE-Tomek Links* memiliki skor akurasi keseluruhan tertinggi diperoleh sebesar 99,02% dan hasil akurasi kelas minoritas *recall* meningkat sebesar 20% dibandingkan dengan data yang tidak seimbang (Rai dkk, 2022).

Algoritma *machine learning* yang diterapkan untuk klasifikasi sentimen pada penelitian sebelumnya, yaitu *Random Forest* (RF), *Bagging*, *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Naïve Bayes* (NB), dan *Decision Tree*. Metode algoritma *Support Vector Machine* (SVM) dipilih sebagai metode klasifikasi sentimen dalam penelitian ini karena *Support Vector Machine* (SVM) bekerja dengan mencari *hyperplane* sebagai cara memaksimalkan jarak pemisah antar kelas (Sain dan Purnami, 2015). Metode *Support Vector Machine* dipilih karena dalam proses penelitian ini menggunakan data ulasan konsumen (Lee dkk, 2022). Jenis data ulasan konsumen merupakan data teks informal yang memiliki banyak *noise* dan tidak mengindahkan aturan tata bahasa (Li dkk, 2019). *Support Vector Machine* sangat kritis dalam proses pengolahan data dan cenderung menghasilkan performa akurasi yang lebih baik dibandingkan metode pada penelitian sebelumnya (Hadwan dkk, 2022).

Tantangan utama dalam klasifikasi *machine learning* adalah menangani data yang tidak seimbang karena data yang tidak seimbang dapat menghasilkan bias hasil terhadap kelas mayoritas dan kinerja klasifikasi yang buruk (Bin Alias dkk, 2021). Data yang tidak seimbang terjadi karena keseimbangan yang tidak merata dalam jumlah pola positif dan negatif atau label kelas *biner* dalam kumpulan data (Sasada dkk, 2020). Salah satu metode *oversampling* adalah teknik *synthetic minority oversampling* (SMOTE) untuk klasifikasi data tidak seimbang menggunakan analisis pohon keputusan. Penerapan metode *undersampling*, masalah *borderline* dan *noise* akan dideteksi oleh *Tomek Links*. *Tomek Links* bekerja dengan cara menghapus kelas negatif dan juga kelas positif yang memiliki karakteristik serupa (Sain dan Purnami, 2015).

Hasil uraian di atas menunjukkan bahwa data ulasan konsumen dapat memberikan informasi yang berguna untuk melacak sentimen, mengukur minat dan kekhawatiran konsumen serta melacak tingkat minat konsumen. Hal ini dikarenakan konsumen dapat lebih jujur dalam mengutarakan pendapatnya melalui ulasan konsumen yang disediakan *platform* pasar *online*. Hasil ulasan yang telah melalui *text processing* dilanjutkan dengan memvisualisasikan kata-kata yang muncul pada sentimen positif dan sentimen negatif dengan menggunakan word cloud. *Word cloud* adalah metode *text mining* yang menampilkan grafis dari frekuensi kata yang lebih menonjolkan kata-kata yang lebih sering muncul dalam teks sumber. Semakin besar ukuran kata dalam visual, maka semakin umum kata tersebut dalam dokumen (Alamoodi dkk, 2021).

Penelitian ini menerapkan analisis sentimen tingkat dokumen melalui ulasan yang diberikan pada kolom ulasan di *marketplace* "<https://shopee.co.id/>" terkait produk *fashion* yang perusahaan gunakan, ulasan konsumen mampu mengantarkan pembaca ke arah ketertarikan konsumen untuk meningkatkan pendapatan perusahaan. Metode sentimen analisis yang digunakan pada penelitian ini adalah SVM dengan kombinasi SMOTE dan *Tomek Links*. SMOTE dan *Tomek Links* untuk memperbaiki alur kerja data *sampling* (Bin Alias dkk, 2021). SVM memiliki kelebihan dalam menentukan jarak pemisah antar kelas dengan menggunakan *support vector* sehingga proses komputasi yang dilakukan pada data yang cukup

besar bisa menjadi lebih cepat (Vapnik dan Cortes 1995). *Support Vector Machine* dipilih karena tingkat keakurasian yang cukup tinggi, sekitar 80 hingga 90 persen, serta pengimplementasian algoritma yang cukup mudah dan fleksibilitas algoritma yang dapat digabungkan dengan metode lainnya (Borg dan Boldt, 2020).

1.2 Tujuan Penelitian

Tujuan penelitian ini adalah mengidentifikasi ulasan konsumen produk *fashion* yang memengaruhi hasil analisis sentimen berdasarkan hasil *word cloud*. Tujuan berikutnya adalah memberikan rekomendasi sebagai hasil untuk mendukung perkembangan bisnis *online* dengan menggunakan metode SVM dan kombinasi SMOTE dengan *Tomek Links* untuk meningkatkan performa model klasifikasi SVM.

1.3 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini sebagai acuan perencanaan dan perancangan strategi pelaku bisnis *online* untuk meningkatkan pendapatan perusahaan. Bagi penulis penelitian ini memberikan manfaat yang cukup besar dalam memahami pengaplikasian penerapan keilmuan *Business Intelligent*.

Sekolah Pascasarjana