

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ulasan palsu yang biasa dikenal sebagai *spammer* sering diklasifikasikan sebagai opini *spam*, opini menipu atau ulasan *spam*. Ulasan palsu dapat menyebabkan masalah keuangan bagi penyedia layanan dan manufaktur produk karena efek buruk dari ulasan palsu untuk reputasi suatu merek bisnis. Suatu bisnis juga dapat kehilangan pelanggan karena ulasan palsu yang dapat memberikan keuntungan yang tidak semestinya kepada pesaing bisnis lain (Mohawesh dkk, 2021).

Dalam menghadapi tantangan ini, perlu adanya sistem yang dapat mendeteksi ulasan palsu. Namun, ulasan palsu akan sulit dibedakan hanya dengan interpretasi manusia. Untuk mendeteksi ulasan palsu dibutuhkan sistem berdasarkan algoritma *machine learning* seperti *text mining* yang telah banyak dibuat dan dikembangkan. Karena ulasan palsu berdampak negatif pada bisnis, maka diperlukan alat yang efektif yang dapat mendeteksi ulasan palsu dengan tingkat keberhasilan yang tinggi. Untuk dapat mendeteksi apakah suatu ulasan itu adalah asli atau palsu diantara sejumlah besar data merupakan hal yang sangat sulit bagi individu atau perusahaan. Algoritma *machine learning* dapat mengatasi kesulitan ini melalui kemampuan untuk mengidentifikasi pola tersembunyi dalam data yang tidak dapat diidentifikasi manusia (Birim dkk, 2022).

Pada *text mining*, teknik yang paling populer yaitu teknik analisis sentimen. Secara khusus, analisis sentimen dalam ulasan produk adalah proses mengeksplorasi ulasan untuk menentukan keseluruhan pendapat atau perasaan tentang suatu produk (Kauffmann dkk, 2020). Dalam penjualan produk secara *online*, tujuan dari analisis sentimen adalah untuk memahami perasaan seseorang tentang produk atau merek tertentu. Analisis sentimen dilakukan berdasarkan asumsi bahwa seseorang secara eksplisit mengungkapkan pendapat mereka tentang merek atau produk dalam teks yang mereka tulis (Birim dkk, 2022). Salah satu pendekatan dalam analisis sentimen yaitu pendekatan berbasis leksikon.

Pendekatan berbasis leksikon menggunakan leksikon sentimen untuk mengumpulkan informasi tentang kata dan frasa yang memiliki nilai sentimen. Pada teknik ini, yang dilakukan terlebih dahulu yaitu membuat leksikon sentimen melalui penyusunan daftar kata sentimen seperti pendekatan secara manual, pendekatan leksikal, dan pendekatan berbasis korpus. Kemudian, menentukan skor polaritas ulasan yang diberikan berdasarkan indikator positif dan negatif yang diidentifikasi dalam leksikon (Bonta dkk, 2019).

Pada penelitian sebelumnya, Birim, dkk (2022) menggunakan analisis sentimen berbasis leksikon yaitu VADER untuk mengidentifikasi skor polaritas dari setiap ulasan pada kumpulan data ulasan produk palsu Amazon. Namun, pada penelitian Birim, dkk, akurasi yang didapatkan menggunakan VADER masih rendah jika dibandingkan dengan pendekatan lain. Kauffmann, dkk (2020) menggunakan analisis sentimen AFINN yaitu analisis sentimen berbasis leksikon untuk membantu mengukur sentimen pengulas. Skor yang diperoleh dari analisis sentimen akan mewakili evaluasi berbagai karakteristik produk. Dengan demikian, leksikon yang berbeda dapat memberikan hasil yang berbeda, tetapi Kauffmann, dkk mendapatkan adanya kekurangan yaitu berupa adanya duplikasi ulasan yang menyebabkan ulasan harus diseleksi dengan menentukan ambang batas pada ulasan.

Pada penelitian ini, analisis sentimen berbasis leksikon yaitu *Textblob* diterapkan untuk menentukan label sentimen dan menghitung skor polaritas pada setiap ulasan. Alasan pemilihan *Textblob* karena pustaka *Textblob* bisa menjadi pustaka yang lebih cepat dan ringkas dalam proses analisis sentimen. *Textblob* juga memungkinkan analisis sentimen teks yang tepat dan memungkinkan terjemahan teks dari suatu bahasa ke bahasa lain (Aljedaani dkk, 2022). *Textblob* merupakan algoritma pembelajaran tanpa pengawasan (*unsupervised learning*) yang dimana tahap analisisnya tanpa memerlukan kumpulan data yang berlabel. *Textblob* dapat memberikan label terhadap dokumen berdasarkan positif atau negatif (Batista dkk, 2021).

Vektorisasi pada teks dapat menggunakan teknik yang disebut *CountVectorizer*. Menurut Daiv, dkk (2020) jika dibandingkan dengan metode

vektorisasi lainnya, *CountVectorizer* memiliki kemampuan untuk memberikan hasil yang lebih akurat pada deteksi ulasan palsu. Penelitian yang dilakukan Shah dan Ahmed (2019) juga menggunakan *CountVectorizer* untuk penelitian deteksi ulasan palsu pada *e-commerce* dengan menerapkan *Principal Component Analysis*. Birim, dkk (2022) menerapkan *CountVectorizer* untuk membangun distribusi frekuensi kata pada setiap ulasan pada kumpulan data ulasan produk palsu Amazon. *CountVectorizer* didasarkan pada jumlah kemunculan kata dalam dokumen yang dimana metode ini mengubah kumpulan dokumen teks menjadi matriks jumlah token (Tripathy, dkk, 2018). Dari hasil penelitian sebelumnya, pada penelitian ini, *CountVectorizer* digunakan untuk menghitung frekuensi kemunculan kata dan melakukan vektorisasi dalam ulasan.

Dalam analisis deteksi ulasan palsu juga dibutuhkan algoritma klasifikasi *machine learning*. Zhang, dkk (2022) menerapkan *K-Nearest Neighbor* (KNN) pada deteksi ulasan palsu untuk menangani ketidakseimbangan data berdasarkan algoritma LDA dan *Kullback-Leibler Divergence*. Hasil yang didapatkan mampu mengungguli semua metode untuk klasifikasi data tidak seimbang pada deteksi ulasan palsu. Menurut Zhang, dkk (2022) Algoritma *K-Nearest Neighbor* dapat mengklasifikasikan pengulas yang tidak memiliki label dalam kumpulan data pengujian untuk mendeteksi pengulas yang curang.

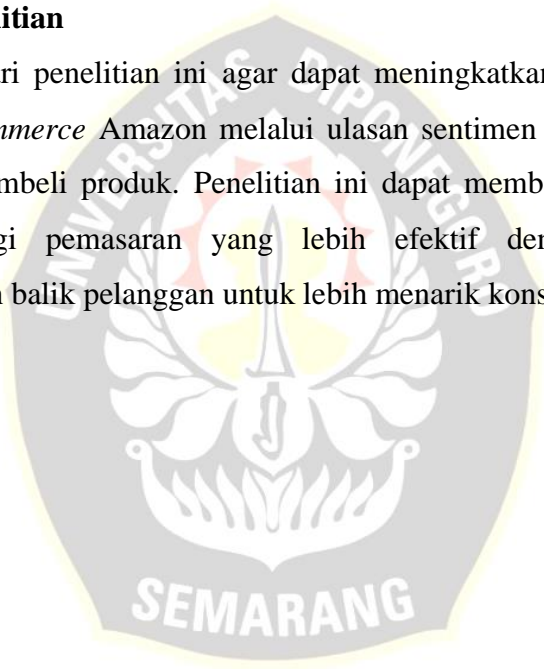
Dalam menentukan nilai parameter yang optimal pada algoritma klasifikasi, salah satu algoritma yang dapat digunakan yaitu *Particle Swarm Optimization* (PSO). Posisi partikel dalam algoritma PSO sesuai dengan kandidat solusi untuk masalah optimisasi (He dkk, 2022). Aburomman dan Reaz (2016) menggunakan SVM dan KNN dengan algoritma optimasi PSO untuk membuat kombinasi klasifikasi dengan akurasi yang lebih baik untuk deteksi intrusi. Rahayu, dkk (2022) menggunakan KNN dan PSO untuk meningkatkan akurasi hasil klasifikasi pada penelitian analisis sentimen. Hasil yang didapat menunjukkan bahwa kombinasi KNN dan PSO mampu meningkatkan nilai akurasi klasifikasi. Pada penelitian ini, klasifikasi KNN dan algoritma optimasi PSO diterapkan untuk meningkatkan kinerja dari algoritma klasifikasi dan meningkatkan akurasi yang lebih baik dalam deteksi ulasan palsu.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini untuk mengeksplorasi peran sentimen dalam ulasan produk Amazon dengan menggunakan algoritma *K-Nearest Neighbor* untuk mendapatkan akurasi terbaik dari parameter k yang ditetapkan kemudian nilai akurasi terbaik dioptimasi menggunakan algoritma *Particle Swarm Optimization* untuk memperoleh hasil kinerja terbaik dalam analisis sentimen dalam ulasan produk Amazon.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini agar dapat meningkatkan nilai jual terhadap produk pada *e-commerce* Amazon melalui ulasan sentimen yang diberikan oleh pengulas yang membeli produk. Penelitian ini dapat membantu Amazon dalam merancang strategi pemasaran yang lebih efektif dengan menyesuaikan berdasarkan umpan balik pelanggan untuk lebih menarik konsumen potensial.



SEKOLAH PASCASARJANA