

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*E-commerce* merupakan suatu set dinamis teknologi, aplikasi dan proses bisnisnya yang menghubungkan perusahaan, konsumen dan komunitas tertentu melalui transaksi elektronik dan perdagangan barang, pelayanan, dan informasi yang dilakukan secara elektronik [1]. Salah satu layanan penyedia *e-commerce* adalah Amazon. Amazon.com adalah situs layanan belanja *online* yang populer di Amerika Serikat. Amazon.com menjual berbagai macam produk yang dibutuhkan masyarakat mulai dari produk kebutuhan sehari-hari hingga produk layanan internet seperti layanan *web*, layanan penyimpanan data dan layanan *cloud computing* [2]. Masyarakat Indonesia juga sering menggunakan layanan Amazon untuk belanja *online*. Untuk meningkatkan layanan pelanggan, maka Amazon.com harus memastikan bahwa barang atau jasa yang terjual oleh Amazon berkualitas baik. Oleh karena itu diperlukan sebuah sarana bagi pembeli untuk menilai produk-produk yang dijual oleh penjual yang terdaftar di Amazon. *Online customer review* adalah ulasan yang diberikan oleh konsumen terkait dengan evaluasi suatu produk dalam berbagai macam aspek, dengan adanya informasi ini konsumen bisa mengetahui kualitas dari produk yang dicari lewat ulasan dan pengalaman yang ditulis oleh konsumen yang telah membeli produk tersebut secara *online* [3]. Biasanya konsumen akan mencari beberapa informasi yang berkualitas sebelum memutuskan untuk membeli suatu produk. Sehingga memberikan keuntungan pada konsumen, yaitu konsumen tidak harus mengunjungi penjual yang berbeda secara langsung dalam menemukan informasi [3]. Dengan begitu, penelitian dilakukan untuk memberikan 2 manfaat yaitu bagi penjual produk di Amazon, yaitu agar penjual barang atau jasa bisa meningkatkan layanannya dengan memilah ulasan positif dan negatif sehingga layanan penjual-penjual tersebut bisa lebih bermanfaat bagi pelanggan dan bagi pembeli, dapat mengambil keputusan dengan cepat saat menentukan produk mana yang harus dibeli karena dapat memilah ulasan-ulasan positif dan negatif dengan cepat. Untuk memahami *review* dari pelanggan dapat dilakukan dengan *sentiment analysis* atau analisis sentimen. *Sentiment analysis* atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen

yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek yang dikemukakan oleh seseorang, apakah cenderung berpandangan negatif atau positif [4]. Analisis sentimen adalah bagian dari *Natural Language Processing* (NLP). NLP adalah cabang ilmu kecerdasan buatan yang mempelajari komunikasi antara manusia dan komputer [5]. Dengan begitu, komputer dapat memproses bahasa manusia dengan baik. Pada penelitian ini dilakukan analisis sentimen terhadap dataset *customer review* dari Amazon dengan melakukan perbandingan 2 algoritma yaitu *Naive Bayes* dan *Logistic Regression*. Setelah itu, model akan dievaluasi dengan *confusion matrix* dengan mempertimbangkan metrik-metrik yaitu *presicion*, *accuration*, *recall* dan *f1-score*.

Ada beberapa penelitian yang menjadi acuan untuk penelitian ini. Diantaranya, yang pertama pada penelitian yang dilakukan oleh Widodo dkk., dilakukan klasifikasi untuk menentukan apakah seseorang terkena diabetes atau tidak dengan membandingkan 4 algoritma yaitu *K-Nearest Neighbor*, *J48*, *Naive Bayes* dan *Logistic Regression*. Hasil dari penelitian ini menunjukkan bahwa model *K-Nearest Neighbor* memiliki hasil akurasi tertinggi yaitu 98%, model *J-48* memiliki akurasi sebesar 95.5%, model *Linear Regression* memiliki akurasi sebesar 92.3 % dan terkecil model *Naive Bayes* memiliki akurasi sebesar 87.11% [6].

Yang kedua, pada penelitian yang dilakukan oleh Kardila, dilakukan sentimen analisis untuk mengklasifikasi *review* terhadap film-film yang ada di IMDB menggunakan algoritma *Naive Bayes*. Diperoleh akurasi sebesar 89.18% dengan perbandingan data *train* dan data *test* sebesar 70 : 30 dan untuk mengambil porsi data *test* terbaik, digunakan *K-Fold* sebanyak 10 kali iterasi dan ditemukan akurasi terbaik saat pada pengujian ke 8 (k-8). Dan rata-rata akurasi dari 10 pengujian tersebut adalah 88.42% [7].

Pada penelitian ini akan dilakukan pemodelan analisis sentimen dengan membandingkan algoritma *Logistic Regression* dan *Multinomial Naive Bayes* dengan 2 rasio pembagian data *train* dan data *test* yaitu 80:20 dan 70:30, setelah itu dilakukan *Synthetic Minority Oversampling Technique* (SMOTE) dan model akan dilatih lagi dengan skenario-skenario yang sama seperti sebelumnya, tetapi dibandingkan hasilnya dengan model yang menggunakan data *train* yang sudah di SMOTE untuk mengetahui apakah SMOTE berdampak dalam meningkatkan hasil akurasi klasifikasi.

## 1.2 Rumusan Masalah

Rumusan masalah yang dipakai dalam penelitian ini adalah sebagai berikut.

1. Bagaimana membuat model analisis sentimen untuk menentukan apakah sebuah ulasan produk di Amazon adalah positif atau negatif menggunakan algoritma *Logistic Regression* dan *Multinomial Naive Bayes*?
2. Bagaimana hasil evaluasi model dengan metrik *presicion*, *recall*, *accuration* dan *f1-score* berdasarkan *confusion matrix* pada model *Logistic Regression* dan *Multinomial Naive Bayes*?
3. Bagaimana pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) terhadap hasil akurasi klasifikasi algoritma *Logistic Regression* dan *Multinomial Naive Bayes*?

## 1.3 Batasan Permasalahan

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Dataset yang diambil berasal dari Kaggle dan mempunyai 11 kolom dan 4915 baris data. Namun kolom yang ditampilkan pada Google Colab hanya 4 yaitu kolom *reviewerName* yang berisi nama-nama dari orang-orang yang memberi ulasan produk, *overall* yang berisi *rating* yang berskala 1-5 dari setiap pelanggan, *reviewText* yang berisi ulasan dari masing-masing pelanggan dan *reviewTime* yang berisi tanggal-tanggal saat ulasan-ulasan tersebut dibuat, variabel independen yang dipakai adalah kolom *reviewText*.
2. *Dataset* yang diambil berbahasa Inggris.
3. Algoritma yang digunakan dalam penelitian ini adalah *Logistic Regression* dan *Multinomial Naive Bayes*.
4. *Labelling* dilakukan secara otomatis dengan bantuan *library* *TextBlob* dari Python.
5. Hasil dari penelitian ini hanyalah model saja.

#### 1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Membuat model analisis sentimen untuk menentukan apakah sebuah ulasan produk di Amazon adalah positif atau negatif menggunakan algoritma *Logistic Regression* dan *Multinomial Naive Bayes*.
2. Mengetahui hasil evaluasi model dengan metrik *presicion*, *recall*, *accuration* dan *f1-score* berdasarkan *confusion matrix* pada model *Logistic Regression* dan *Multinomial Naive Bayes*.
3. Mengetahui pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) terhadap hasil akurasi klasifikasi algoritma *Logistic Regression* dan *Multinomial Naive Bayes*.

#### 1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah sebagai berikut.

1. Bagi Penjual Amazon, dapat mengetahui apakah produk yang dijual baik atau buruk berdasarkan *review* dari pelanggan sehingga dapat menjadi bahan pertimbangan dalam konsistensi dan peningkatan layanan barang atau jasa untuk memuaskan pelanggan.
2. Bagi pembeli Amazon, dapat mengetahui apakah *review* dari pembeli lain mengenai sebuah produk itu positif atau negatif sehingga bisa menjadi bahan pertimbangan dalam membeli produk.
3. Bagi mahasiswa dan peneliti lainnya, dapat menjadi pertimbangan untuk penelitian selanjutnya.

#### 1.6 Sistematika Penulisan

Sistematika penulisan laporan ini adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab 1 berisi pemaparan awal penelitian yaitu latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan laporan.

- Bab 2 LANDASAN TEORI

Bab 2 berisi kumpulan-kumpulan literatur yang akan dipakai dalam penelitian yaitu *Amazon*, *machine learning*, *Natural Language Processing*, metode *preprocessing*, metode ekstraksi fitur, *Synthetic Minority Oversampling Technique (SMOTE)*, *Naive Bayes*, *Logistic Regression* dan *confusion matrix*.

- Bab 3 METODOLOGI PENELITIAN

Bab 3 berisi gambaran umum penelitian dan perancangan sistem secara detail.

- Bab 4 HASIL DAN DISKUSI

Bab 4 berisi spesifikasi sistem yang digunakan untuk penelitian dan pembuatan sistem yang sudah dirancang sebelumnya.

- Bab 5 KESIMPULAN DAN SARAN

Bab 5 berisi kesimpulan yang diperoleh selama penelitian dan saran untuk penelitian selanjutnya.

