



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

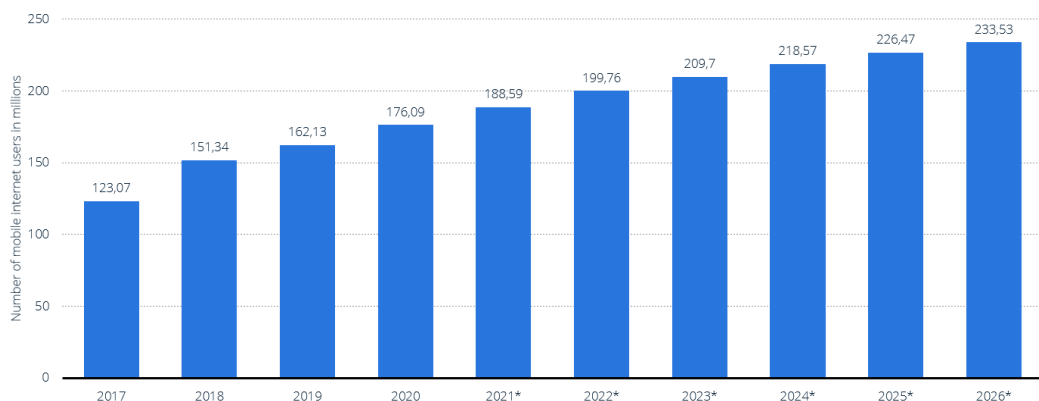
This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

E-commerce atau perdagangan secara elektronik adalah suatu kegiatan jual beli barang ataupun jasa melalui internet. Toko online bukanlah hal yang asing lagi untuk didengar di masa kini. Kegiatan perbelanjaan dan jual beli barang sudah sangat mudah berkat kemajuan teknologi yang sangat pesat. Kita tidak perlu pergi berbelanja ke pasar lagi untuk membeli barang dan kebutuhan sehari-hari, cukup dengan ponsel cerdas dan koneksi ke internet, kita dapat berbelanja dengan mudah kapan dan dimanapun kita berada [1].

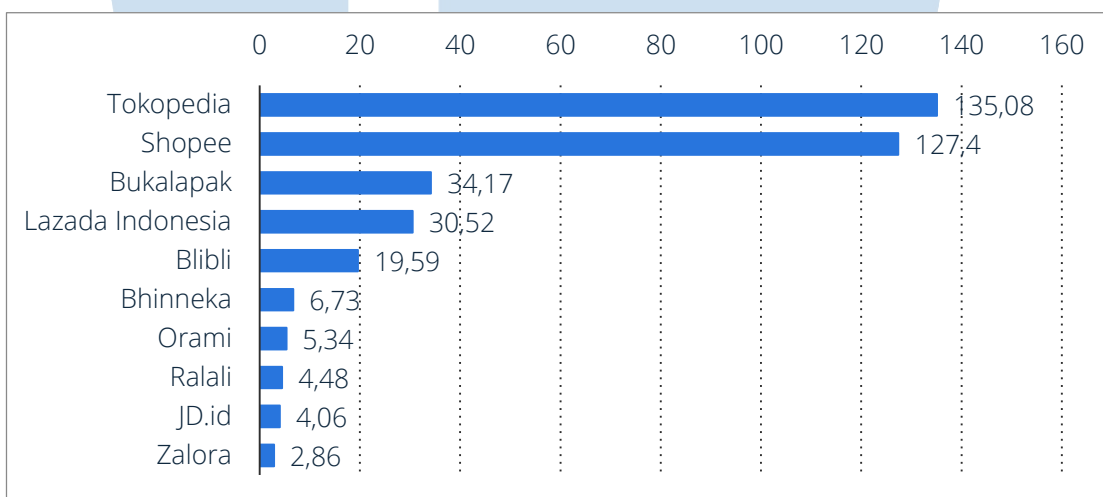


Gambar 1. 1 Pengguna Internet *Mobile* tahun 2017-2026

Gambar 1.1 adalah total pengguna internet mobile di Indonesia meningkat sangat signifikan dari tahun ke tahun juga didukung dengan perkembangan teknologi yang sangat pesat dan juga infrastruktur di Indonesia. Data Asosiasi Penyelenggara Jasa Internet Indonesia mencatat penetrasi pengguna Internet di

Indonesia mencapai 64,8% dengan rata-rata penggunaan media sosial harian mencapai 180 menit pada tahun 2018 .

Melihat pertumbuhan pengguna internet mobile yang semakin bertumbuh, hal ini juga terjadi pada pengguna layanan e-commerce. Menurut data dari hasil survey *We Are Social* pada April 2021 bahwa 88.1% dari pengguna internet di Indonesia menggunakan layanan e-commerce. Dimana Indonesia menduduki peringkat tertinggi di dunia pengguna internet yang menggunakan layanan e-commerce [2].



Gambar 1. 2 Top 10 Situs e-commerce di Indonesia Q1 2021  
Sumber: Statista (2021)

Gambar 1.1 merupakan grafik jumlah pengunjung 10 *e-commerce* teratas di Indonesia pada kuartal ke-1 tahun 2021. Data tersebut menunjukkan bahwa Tokopedia, Shopee, dan Bukalapak menempati posisi teratas dari e-commerce lainnya. Tokopedia memiliki pengunjung sebanyak 135.08 juta pengunjung, Shopee memiliki pengunjung sebanyak 127.4 juta pengunjung, dan Bukalapak memiliki pengujung sebanyak 34.17 juta pengunjung [3].

Masyarakat dapat memberikan pendapat dan opini mereka secara bebas melalui media sosial *Twitter* salah satunya. Masyarakat memilih *Twitter* dikarenakan *Twitter* termasuk cepat dalam menyebarkan dan mengabarkan melalui postingan *Twitter* dan juga bisa cepat viral jika banyak yang *re-tweet* dan menyukai *tweet*. Menurut data kominfo, Indonesia menjadi negara kelima terbesar pengguna *Twitter* dibawah Inggris dan negara besar lainnya, dari 245 juta penduduk Indonesia, pengguna *Twitter* di Indonesia mencapai 19.5 juta [4].

Namun dengan banyaknya opini yang ditumpahkan para pengguna *Twitter* tersebut tidaklah selalu benar, ada kalanya suatu pihak memberikan opini yang palsu. Seperti memberikan ulasan sendiri yang positif untuk menaikkan nama toko atau produk, namun juga bisa kompetitor memberikan ulasan palsu berupa ulasan negatif terhadap toko atau produk kompetitor. Hal ini tentu membuat calon konsumen bingung bagaimana cara melihat ulasan tersebut palsu atau asli. Karena ulasan pengguna seharusnya ditujukan untuk memudahkan calon konsumen untuk melihat bagaimana kualitas barang atau layanan dari toko online tersebut.

Pengambilan data dilakukan menggunakan metode *text mining* dengan aplikasi RapidMiner. *Twitter* telah menyediakan token khusus untuk kepentingan analisis yang bisa mendapatkan akses mengambil data *Twitter* secara langsung. Data yang penulis kumpulkan berjumlah data kotor dari setiap toko online yang ada, yaitu Tokopedia 15.000 *tweet*, Shopee 15.000 *tweet*, dan Bukalapak 15.000 *tweet* pada bulan September sampai November tahun 2021.

Analisis sentimen memegang peranan yang penting dalam mengidentifikasi sentimen pengguna toko online tentang produk dan layanannya [5]. Analisis

sentiment dapat menganalisis pendapat, sentimen, evaluasi, penilaian, pendirian, dan emosi terhadap suatu entitas seperti produk, layanan, dan masalah [6]. Karena opini atau pendapat sangat penting dalam membuat keputusan baik dari sisi penjual maupun calon pembeli. Sehingga dilakukan analisis sentimen untuk deteksi ulasan palsu terhadap toko online.

Analisis sentiment ini menggunakan tiga algoritma pembandingan yaitu *Naïve Bayes*, *Support Vector Machine*, dan *Logistic Regression*. *Naïve Bayes* menggunakan fitur *bag of words* dalam mengidentifikasi penipuan dalam klasifikasi teks, membagi dataset menjadi data latih dan data uji kemudian probabilitas kata-kata akan dihitung dengan prediksi teorema *Bayes*. *Support Vector Machine* melatih model untuk mendeteksi ulasan palsu dengan mengambil matriks dan label. Label yang menunjukkan apakah ulasan itu palsu atau tidak. SVM banyak digunakan dalam deteksi ulasan palsu karena sering menghasilkan akurasi yang tinggi. *Logistic Regression* menggunakan fungsi logistic untuk membangun variabel biner untuk mengetahui probabilitas kelas tertentu seperti palsu atau nyata. Ketiga algoritma ini digunakan karena memiliki akurasi yang lebih tinggi dalam klasifikasi ulasan palsu [5]. Dibandingkan dengan algoritma *Decision Tree*, *K-Star*, dan *KNN* yang memiliki akurasi yang cukup atau lebih rendah dibandingkan *Support Vector Machine*, *Naïve Bayes*, dan *Logistic Regression* [7].

Oleh karena itu, untuk mengolah informasi yang diambil dari *Twitter* tersebut, maka dilakukan pengolahan data dengan analisis sentimen, yaitu dengan memisahkan pendapat dan opini masyarakat ke dalam kelas positif asli, kelas

negatif asli, kelas positif palsu, dan kelas negatif palsu. Pada akhirnya akan diambil kesimpulan berupa faktor apa yang sering disebut atau dibahas di dalam pendapat dan opini tersebut. Analisis sentimen terhadap *Twitter* seringkali dilakukan dengan algoritma *Naïve Bayes classification*, *Support Vector Machine (SVM)* dan *Logistic Regression*. Pada penelitian ini akan dilakukan perbandingan hasil dari algoritma tersebut untuk mengetahui algoritma mana yang memiliki akurasi yang lebih tinggi.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas dapat disimpulkan permasalahan yang ada yaitu:

1. Bagaimana hasil *text mining* data *Twitter*?
2. Algoritma apa yang menghasilkan akurasi paling tinggi dalam deteksi ulasan palsu terhadap toko online?
3. Bagaimana hasil deteksi ulasan palsu pada toko online Tokopedia, Shopee, dan Bukalapak?
4. Bagaimana hasil perbandingan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Logistic Regression*?

## 1.3 Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data berupa teks yang diambil dari media sosial *Twitter* bulan November 2021 sebanyak 15.000 *tweet* setiap kata kunci dengan spesifik kata kunci toko online seperti: Tokopedia, Shopee, Bukalapak.

2. Pemberian label data *tweet* terdapat empat kelas tanpa kelas netral yaitu *true positive, true negative, fake positive, dan fake negative*.
3. Hasil akhir dari penelitian ini menunjukkan seberapa besar akurasi dan total data kelas positif dan negatif dari ulasan asli dan palsu terhadap toko online menggunakan algoritma *Naïve Bayes, Support Vector Machine, dan Logistic Regression*.

## **1.4 Tujuan dan Manfaat Penelitian**

### **1.4.1 Tujuan Penelitian**

Tujuan yang akan dicapai yaitu mendapatkan *tweet* dari proses *text mining*, mengetahui persentase ulasan palsu serta mengetahui ulasan tersebut termasuk kedalam kategori positif atau negatif terhadap toko online dari *tweet* media sosial *Twitter*, hasil perbandingan klasifikasi dari algoritma *Naïve Bayes, Support Vector Machine, Logistic Regression*, dan menemukan algoritma terbaik yang memiliki akurasi tertinggi untuk klasifikasi sentiment ulasan palsu dari media sosial.

### **1.4.2 Manfaat Penelitian**

Setelah mendapatkan hasil dari analisis sentimen, maka hasil dari penelitian ini dapat menjadi informasi kepada pengguna dan juga toko online agar dapat lebih mengerti dan memilah ulasan-ulasan yang beredar. Selain itu, hasil penelitian ini juga dapat dijadikan acuan atau sumber referensi bagi penelitian yang sejenis atau penelitian selanjutnya.