

# BAB I

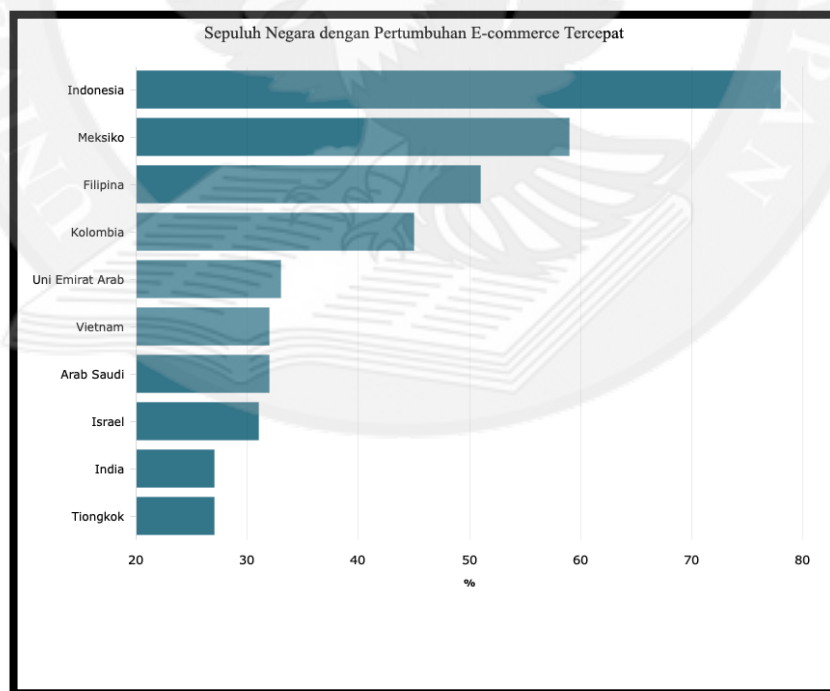
## PENDAHULUAN

### 1.1 Latar Belakang

Indonesia sebagai salah satu bangsa yang hidup dalam zaman perkembangan teknologi informasi yang sangat maju, secara tidak langsung juga harus terlibat dalam pesatnya perkembangan teknologi saat ini untuk kepentingan bangsa sendiri juga. Kehidupan saat ini, dengan adanya teknologi Internet, sangat membantu kehidupan manusia, mulai dari bidang pendidikan, kesehatan maupun bisnis. Di zaman sekarang, teknologi Internet dibutuhkan sebagai sarana melakukan komunikasi dan memfasilitasi pengelolaan arus informasi hampir dalam setiap aspek kehidupan, salah satunya dapat berupa transaksi berjalannya suatu bisnis [1]. Internet juga menjadi sarana bagi masyarakat untuk mendapatkan informasi, apalagi dengan munculnya media sosial. Pengguna Internet dan media sosial yang banyak dan terus bertambah ini pun memicu munculnya peluang bisnis baru di Indonesia, salah satunya yaitu munculnya perusahaan *marketplace* di Indonesia.

*Online marketplace* merupakan tempat yang memfasilitasi jual beli secara *online* yang dikelola oleh satu pihak tetapi produk dan informasinya bisa disediakan oleh produsen lain sebagai pihak ketiga [2]. *Online marketplace* memiliki konsep dasar seperti pasar tradisional, bedanya yaitu *online marketplace* ini merupakan versi digital dari pasar tradisional. *Online marketplace* menjadi tempat bertemunya penjual dan pembeli secara digital atau daring. Di Indonesia saat ini pun sudah mulai banyak perusahaan *marketplace* contohnya seperti Tokopedia, Shopee, Bukalapak dan lainnya. Kehadiran *online marketplace* tersebut membuat masyarakat mempunyai pilihan atas *online marketplace* sesuai preferensinya. Salah satu faktor yang menjadi dasar pemilihan tersebut adalah dengan membaca komentar atau *review* dari konsumen atas *marketplace* yang diunggah di media sosial.

Majunya perkembangan teknologi saat ini, konsumen tentu dapat melihat bagaimana pandangan masyarakat mengenai suatu *marketplace* tertentu, salah satunya yaitu melalui media sosial. Saat ini, media sosial twitter yang telah berdiri sejak tahun 2006 dan terus berkembang seiring berjalannya waktu hingga saat ini menjadi salah satu dari sekian banyaknya media sosial populer di Indonesia. Salah satu dampak yang dirasakan dengan pengguna aktif media sosial yang dapat dikatakan sangat banyak setiap harinya adalah bertambah banyaknya *review*, ulasan, dan komentar pada media sosial. Hal tersebut memungkinkan karena betapa mudahnya masyarakat memberikan *review*, ulasan dan komentar pada media sosial hanya dengan sentuhan jari saja [3]. *Review*, ulasan, serta komentar pada media sosial ini dapat dijadikan suatu referensi bagi masyarakat dalam memilih *marketplace* mana yang ingin digunakan. *Review*, ulasan, serta komentar pada media sosial ini juga dapat mempengaruhi niat beli konsumen apalagi negara Indonesia menduduki peringkat pertama sebagai negara dengan pertumbuhan *e-commerce* tercepat di dunia menurut data dari situs katadata [4].



Gambar 1.1 Sepuluh negara dengan pertumbuhan *e-commerce* tercepat  
Sumber: situs katadata [4]

Berdasarkan uraian latar belakang tersebut, dapat dilihat sentimen masyarakat dan ditentukan faktor-faktor atau topik-topik apa saja yang dianggap penting oleh konsumen menurut pengalamannya melalui *review*, ulasan, serta komentar pada salah satu media sosial paling populer di Indonesia yaitu twitter. Hal tersebut dapat tercapai dengan menggunakan metode *text-mining*. Tujuan dari *text-mining* yaitu untuk mengumpulkan data berupa teks dalam jumlah yang besar serta mengolahnya menjadi suatu informasi yang berguna. Penelitian ini mengambil data *tweets* dari akun media sosial twitter tiga *online marketplace* dengan pengunjung terbesar yaitu Tokopedia, Shopee, dan Bukalapak. Dalam penelitian ini, data yang diolah menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF) untuk pembobotan katanya (*term weighting*) serta menggunakan *lexicon* untuk penilaian sentimen. Metode TF-IDF ini menjadi merupakan metode algoritma pembobotan kata yang paling dikenal, dan akurasi dalam metode TF-IDF dalam mengklasifikasikan artikel ini dapat dikatakan sangat baik, lalu dilanjutkan dengan proses *clustering*. Salah satu tujuan menggunakan metode *clustering* yaitu untuk mengeksplorasi distribusi dan pola data. Suatu *cluster* biasanya akan memiliki karakteristik atau sifat yang serupa dibandingkan dengan *cluster* lainnya. Pada umumnya, *clustering* digunakan untuk melakukan analisis pola-pola data, mengelompokkan, dan membuat keputusan. Saat ini, sudah cukup banyak algoritma dalam melakukan proses *clustering*, salah satu diantaranya *K-means clustering*, yaitu algoritma *clustering* yang digunakan pada penelitian ini. Dengan kemampuan mengelompokkan data dalam jumlah yang cukup besar serta waktu komputasi yang relatif cepat serta efisien, algoritma *K-means* menjadi salah satu algoritma yang cukup terkenal dalam hal *clustering*. Oleh sebab itu, penelitian ini akan membahas tentang penerapan metode *text mining* dengan *k-means clustering* untuk mengetahui sentimen dan topik populer.

## 1.2 Perumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Bagaimana cara melakukan *text mining* menggunakan *software* R Studio.
- 2) Bagaimana cara melakukan pengolahan data menggunakan metode algoritma *K-means clustering*.
- 3) Topik-topik apa saja yang paling banyak dibahas oleh konsumen melalui media sosial dengan proses *text mining* dan menggunakan metode *K-means clustering*.
- 4) Bagaimana sentimen konsumen terhadap *online marketplace* melalui media sosial dengan proses *text mining* dan menggunakan metode *K-means clustering*.

## 1.3 Pembatasan Masalah

Adapun penelitian ini memiliki batasan-batasan meliputi:

- 1) Penelitian ini menggunakan *software* R Studio.
- 2) Penelitian ini menggunakan metode *K-means clustering*.
- 3) Banyak *marketplace* yang dilibatkan dalam penelitian ini sebagai studi kasus berjumlah tiga, yaitu Tokopedia, Shopee serta Bukalapak.
- 4) Data yang digunakan terdiri dari *tweet mention* berbahasa Indonesia dengan jumlah data yang digunakan 1500 *tweet* dari akun twitter utama Tokopedia (@Tokopedia), Shopee (@ShopeeID) serta Bukalapak (@Bukalapak).

## 1.4 Tujuan dan Manfaat Penelitian

Berdasarkan latar belakang yang telah dijabarkan, maka tugas akhir ini memiliki tujuan dan manfaat sebagai berikut:

#### 1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

- 1) Mengetahui cara melakukan *text mining* menggunakan *software* R Studio.
- 2) Mengetahui cara melakukan pengolahan data menggunakan metode algoritma *K-means clustering*.
- 3) Mengidentifikasi topik-topik yang paling banyak dibahas oleh konsumen melalui media sosial dengan proses *text mining* dan menggunakan metode *K-means clustering*.
- 4) Mengidentifikasi sentimen terhadap *online marketplace* oleh konsumen melalui media sosial dengan proses *text mining* dan menggunakan metode *K-means clustering*.

#### 1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

- 1) Hasil penelitian dapat digunakan untuk menggambarkan informasi mengenai topik-topik apa saja yang paling banyak dibahas oleh konsumen terhadap *online marketplace* di Indonesia.
- 2) Hasil penelitian dapat digunakan untuk menggambarkan informasi mengenai bagaimana sentimen konsumen terhadap *online marketplace* di Indonesia.
- 3) Hasil penelitian dapat dijadikan sebagai faktor pendukung dalam pengambilan keputusan (*decision making*) oleh *online marketplace*.

### 1.5 Metodologi Penelitian

Adapun metodologi penelitian yang digunakan dalam penelitian ini adalah:

#### 1.5.1 Authentication

Tahap *authentication* ini merupakan tahap integrasi antara media sosial Twitter dengan *software* R studio. Tahap *authentication* ini memanfaatkan

teknologi API. Dengan menggunakan API (*application programming interface*), dapat mengintegrasikan dua bagian dari aplikasi atau dengan aplikasi yang berbeda dalam waktu yang bersamaan. R Studio membutuhkan beberapa kode khusus untuk dapat menggunakan API yang disediakan oleh pihak twitter. Beberapa kode tersebut yaitu API key, API secret, access token, access token secret.

#### 1.5.2 Data Collecting

Tahap *data collecting* ini merupakan tahap pengambilan data *tweet* secara *real time* dari media sosial twitter dan dilakukan dengan menggunakan fungsi '*searchTwitter*' dan akan dimasukkan ke dalam satu *file*.

#### 1.5.3 Text Preprocessing

Proses *text preprocessing* ini terdiri dari beberapa tahap yaitu *case folding*, *stemming*, serta *stopwords removing*. Tahap-tahap tersebut juga dilakukan menggunakan *software* R studio.

#### 1.5.4 Word Cloud Creating

*Word cloud creating* merupakan metode yang digunakan untuk memvisualisasikan data teks, tahap ini menggunakan *package* yang telah disediakan oleh R studio.

#### 1.5.5 Lexicon

Tahap *lexicon* ini merupakan tahap pemberian nilai sentimen pada data *tweets*. Tahap ini menggunakan *software* R studio.

#### 1.5.6 Term Weighting

Pada tahap *term weighting* ini menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). *Term Frequency* (TF) merupakan frekuensi untuk menghitung seberapa sering istilah muncul dalam dokumen saat ini yang sebelumnya diproses atau dihitung. Sedangkan *Inverse Document Frequency*



(IDF) dilakukan dengan menghitung nilai istilah saat ini pada seberapa sering muncul di beberapa dokumen.

#### 1.5.7 Clustering

Pada tahap *clustering* ini dilakukan pengelompokan secara otomatis. Setelah dilakukannya *preprocessing* pada setiap dokumen, diakhiri dengan melakukan proses *clustering* dengan menggunakan algoritma *K-means clustering*.

### 1.6 Sistematika Penulisan

Untuk memberikan gambaran yang lebih jelas mengenai laporan ini, maka materi-materi yang tertera pada laporan tugas akhir ini dikelompokkan menjadi beberapa sub-bab dengan sistematika penyampaian sebagai berikut:

#### BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metodologi, serta sistematika penulisan.

#### BAB II LANDASAN TEORI

Bab ini berisikan teori-teori dasar berupa definisi dan penjelasan yang dikutip dari buku ataupun sumber lainnya serta beberapa *literature review* yang berhubungan dengan penelitian dan penyusunan laporan tugas akhir.

#### BAB III METODOLOGI PENELITIAN

Bab ini berisi penjelasan mengenai metodologi penelitian yang digunakan pada penelitian ini.

#### BAB IV IMPLEMENTASI

Bab ini berisi pembahasan mengenai implementasi prosedur *text mining* untuk menyelesaikan masalah-masalah sesuai dengan batasan yang telah ditentukan.

## BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran dari hasil penelitian yang telah dilakukan.

