

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi khususnya internet saat ini telah berkembang pesat. Internet saat ini telah menjadi bagian penting dalam menjalani aktivitas sehari-hari. Perkembangan teknologi informasi dan komunikasi yang begitu pesat tidak terlepas dari penyedia layanan web yang menyediakan informasi yang sangat beragam. Salah satu informasi yang disajikan dapat berupa data teks. Contoh data teks yang dapat diambil adalah data yang berasal dari *twitter*. *Twitter* sendiri merupakan salah satu media sosial yang sudah akrab dengan masyarakat Indonesia. Berdasarkan informasi yang diperoleh dari TiNewss.com, jumlah pengguna *twitter* di Indonesia pada awal tahun 2022 mencapai 18,45 juta.

Pengguna aktif pada *twitter* akan memberikan informasi atau komentar terbaru tentang hal yang sedang dibicarakan sehingga dapat menjadi *trending topic*. Banyak masyarakat Indonesia yang sering menyampaikan pendapat atau opininya melalui *twitter*. Data tentang opini tersebut akan mudah dikumpulkan dikarenakan banyaknya pengguna aktif yang menyampaikan opininya melalui *twitter* (Novantirani, Sabariah, dan Effendy, 2015).

Salah satu opini yang dapat disampaikan pengguna *twitter* adalah ulasan mengenai film yang sedang tayang di bioskop sekaligus memberikan penilaian pada film tersebut. Hal tersebut sebenarnya dapat dimanfaatkan untuk mengklasifikasikan berbagai ulasan sehingga menghasilkan informasi yang baik. Namun dengan banyaknya ulasan yang masuk dan tersusun acak pada *twitter*, pengguna *twitter* akan kesulitan untuk mendapatkan informasi yang penting dari setiap ulasan pada *twitter* (Fransiska dan Yolanda, 2019).

Banyaknya data ulasan yang masuk pada *twitter*, tentu saja akan sangat memakan waktu yang lama apabila menyalin dengan manual. Teknik yang digunakan untuk mengambil data atau informasi semi terstruktur dari situs atau *website* adalah dengan menggunakan *Web Scraping*. *Web Scraping* akan

melakukan ekstraksi data tertentu saja dari *website* yang dituju. Data hasil *Web Scraping* tersebut akan digunakan untuk analisis (Ayani, Pratiwi, dan Muhardi, 2019).

Cara yang dapat digunakan untuk menghasilkan informasi dari data ulasan tersebut adalah dengan *text mining*. *Text mining* adalah proses ekstraksi pola (informasi yang berguna) dari data yang tidak terstruktur yang kemudian akan diperoleh pola dari data tersebut (Feldman dan Sanger, 2007). Salah satu metode *text mining* yang dapat digunakan adalah menggunakan Asosiasi Teks. Asosiasi Teks dapat digunakan untuk mengetahui hubungan atau keterkaitan antar kata tersebut yang muncul bersamaan dalam sebuah dokumen. Asosiasi Teks akan menghasilkan informasi dan topik yang sering dibicarakan (Santoso dan Nugroho, 2019).

Berdasarkan latar belakang tersebut, implementasi *Web Scraping* sangat tepat digunakan untuk memperoleh data ulasan dari suatu situs atau *website* sebelum dianalisis. Kemudian untuk menghasilkan informasi yang bermanfaat dari data tersebut maka analisis yang dilakukan dapat menggunakan Asosiasi Teks.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang dikemukakan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana implementasi *Web Scraping* dalam pengumpulan data ulasan film pada *twitter*.
2. Bagaimana implementasi Asosiasi Teks dalam mendapatkan informasi yang diperoleh dari hasil asosiasi antar kata yang terbentuk.

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang dikemukakan, tujuan dalam penelitian ini adalah:

1. Mengimplementasikan *Web Scraping* untuk mengumpulkan data ulasan film pada *twitter*.

2. Mengimplementasikan Asosiasi Teks untuk mendapatkan informasi dari hasil asosiasi antar kata yang terbentuk.

#### **1.4 Batasan Masalah**

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan merupakan data ulasan Film KKN di Desa Penari yang ada pada komentar di media sosial *twitter*.
2. Pengambilan komentar pada *twitter* dilakukan dengan menggunakan *Software Jupyter Notebook*.
3. Data ulasan yang diambil merupakan ulasan film yang tayang perdana pada tanggal 30 April 2022.
4. Ulasan yang diambil merupakan ulasan yang menggunakan Bahasa Indonesia.
5. Ulasan dipisahkan menjadi ulasan positif dan ulasan negatif.
6. Seleksi komentar yang termasuk ke dalam ulasan serta pemisahan ulasan positif dan ulasan negatif dilakukan berdasarkan subjektif peneliti. Ulasan positif merupakan tanggapan positif atau yang baik mengenai film tersebut, sedangkan ulasan negatif merupakan tanggapan negatif atau tidak baik mengenai film KKN di Desa Penari.
7. Definisi ulasan dan komentar dibatasi sebagai berikut: komentar merupakan *tweet* yang berisi mengenai informasi penayangan film dan keinginan untuk menonton film. Sedangkan ulasan merupakan tanggapan mengenai Film KKN di Desa Penari.

#### **1.5 Tinjauan Pustaka**

Fransiska dan Yolanda (2019) melakukan penelitian dengan menggunakan *Naïve Bayes Classifier* untuk menganalisis *review* film pada *twitter*. Data awal yang diperoleh pada *twitter* sebanyak 99.521 data, kemudian setelah melalui tahap *preprocessing* data yang tersisa sebanyak 36.106 data. Hasil penelitian menggunakan *Naïve Bayes Classifier* menghasilkan dua sentimen yang terdiri dari sentimen *emotion* dan *sentiment polarity*. Sentimen *emotion* mengklasifikasikan

data menjadi enam yang terdiri dari *emotion, anger, disgust, fear, joy, sadness*, dan *surprise*. Sedangkan sentimen *polarity* mengklasifikasikan data menjadi tiga yaitu sentimen positif, sentimen negatif, dan sentimen netral.

Satriajati, Panuntun, dan Pramana (2021) mengimplementasikan *Web Scraping* untuk mengumpulkan berita kriminal pada masa pandemi *Covid-19*. Situs berita yang digunakan adalah detik.com. Hasil penelitian menghasilkan bahwa *Web Scraping* dapat digunakan untuk mengumpulkan berita. Hasil dari *Web Scraping* tersebut akan digunakan untuk mengetahui tren harian jumlah berita kriminal kemudian dibandingkan dengan tren harian jumlah kasus *Covid-19* di Indonesia.

Fide, Suparti, dan Sudarno (2021) melakukan penelitian untuk mengetahui sentimen pada aplikasi *TikTok* di *Google Play* menggunakan *Support Vector Machine* (SVM) dan Asosiasi. Hasil penelitian menunjukkan bahwa terdapat 1741 atau 54,41% ulasan positif dan 1459 atau 45,59% ulasan negatif. Hasil klasifikasi dan asosiasi ulasan positif menunjukkan bahwa pengguna menyukai *TikTok* versi sekarang yang berisi video-video lucu di beranda. Pengguna juga menyatakan bahwa tidak menyesal mengunduh *TikTok* dan mengajak pembaca juga untuk mengunduh. Sedangkan untuk klasifikasi dan asosiasi ulasan negatif menunjukkan bahwa pengguna memberikan ulasan negatif dengan alasan gagal dalam mendaftar atau masuk karena tidak memenuhi syarat tanggal lahir. Pengguna yang memberikan ulasan negatif biasanya juga memberikan penilaian bintang yang sedikit.

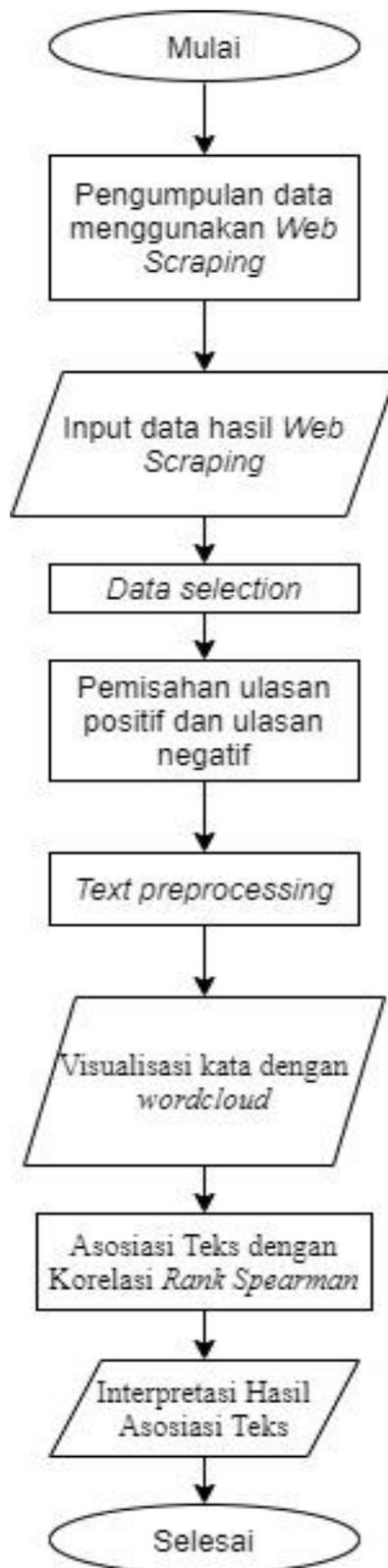
Rahanto dan Kharisudin (2021) melakukan penelitian dengan menganalisis sentimen data ulasan menggunakan *Naïve Bayes Classifier* pada situs *tripadvisor*. Teknik yang digunakan untuk mengambil data pada situs tersebut menggunakan *Web Scraping*. Proses pengklasifikasian selanjutnya adalah dengan menggunakan *Naïve Bayes Classifier*. Hasil klasifikasi menunjukkan bahwa pengunjung lebih banyak memberikan ulasan positif dibandingkan negatif. Hasil klasifikasi menggunakan *Naïve Bayes Classifier* menghasilkan tingkat akurasi sebesar 76,6%. Hasil asosiasi menunjukkan bahwa pengunjung banyak memberikan ulasan positif mengenai kamar hotel yang bergaya modern dan minimalis dengan balkon, staf yang ramah serta membantu, menu restoran yang memiliki banyak variasi

makanan, serta adanya fasilitas penunjang lain seperti taman bermain, kolam renang, *wifi*, dan *jogging track*. Sedangkan hasil asosiasi mengenai ulasan negatif terdiri dari pelayan pembersih kamar kurang teliti saat membersihkan sudut kamar.

## 1.6 Metodologi Penelitian

Metode yang digunakan untuk penelitian ini diawali dengan studi literatur yaitu untuk mencari referensi yang relevan dengan permasalahan yang akan diteliti. Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari media sosial *twitter*. Data yang digunakan merupakan ulasan Film KKN di Desa Penari pada komentar di media sosial *twitter* dari tanggal 30 April 2022. Pengambilan komentar pada *twitter* menggunakan *Web Scraping* dengan menggunakan kata kunci “KKN di Desa Penari”.

Langkah pertama yang dilakukan untuk mendapatkan informasi yang bermanfaat dari komentar *twitter* adalah melakukan *data selection*. *Data selection* merupakan seleksi terhadap komentar yang termasuk ke dalam ulasan dan bukan ulasan. Kemudian dari hasil seleksi dilakukan pemisahan antara ulasan positif dan ulasan negatif. Langkah selanjutnya adalah melakukan *text preprocessing*. Tujuan dari *text preprocessing* adalah untuk memperbaiki informasi dari setiap data ke dalam bentuk yang lebih baku agar dapat dianalisis. Tahapan *preprocessing* terdiri dari *cleansing data*, *case folding*, *spelling normalization*, *filtering*, dan *tokenizing*. Setelah melakukan *text preprocessing*, langkah selanjutnya adalah melakukan visualisasi kata dengan *word cloud* untuk masing-masing ulasan positif dan ulasan negatif. Setelah menemukan kata yang sering digunakan ketika memberikan ulasan dengan melihat visualisasi kata pada *word cloud*, langkah selanjutnya adalah melakukan Asosiasi Teks untuk mengetahui hubungan antar kata yang muncul bersamaan dalam satu dokumen sehingga menghasilkan suatu informasi yang bermanfaat. Asosiasi Teks dilakukan dengan menggunakan korelasi *Rank Spearman*. Langkah terakhir adalah melakukan interpretasi terhadap kata yang terbentuk dan relevan dari hasil Asosiasi Teks. Alur dalam penelitian dapat dilihat pada *flowchart* berikut.



**Gambar 1.1** *Flowchart* Penelitian