

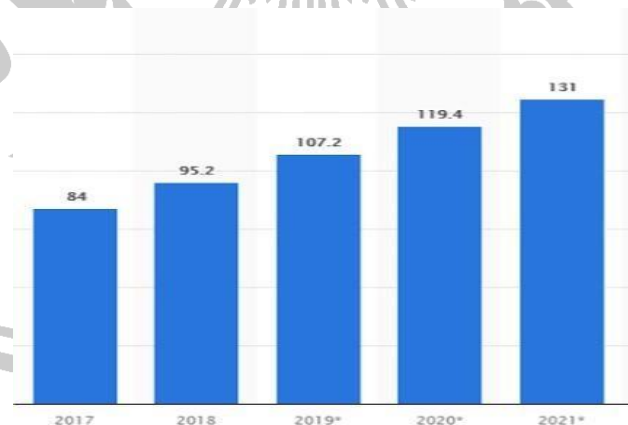
# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan teknologi internet berkembang sangat pesat. Dengan internet, ini memungkinkan pemakainya untuk secara langsung terus selalu berinteraksi dan terhubung dengan pemakai internet lain tanpa dibatasi oleh waktu. Saat ini, inovasi internet telah diterapkan oleh berbagai negara di belahan bumi manapun, salah satunya adalah Indonesia (Alhamda, dkk., 2019). Indonesia adalah salah satu negara dengan pemakai internet yang paling cepat berkembang dengan sekitar 196 juta penduduk atau 73,3% dari populasi terjangkau internet sebagaimana yang dikatakan oleh studi yang dipimpin Afiliasi Pemasok Akses Jaringan Indonesia (APJII) pada tahun 2021. Menurut *Cisco Visual Networking* (VNI) perkembangan teknologi pada tahun 2017 mencapai 3,4 miliar atau 45% dari total populasi di dunia, dan pada tahun 2022 perkembangan teknologi juga diproyeksikan akan menyentuh 4,8 miliar atau sekitar 60% dari populasi di dunia.

**Data Pengguna internet tahun 2017-2022**



Gambar 1. 1 Data Pengguna Internet Dalam Satuan Juta

*Sentiment analysis* adalah proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber. Salah satu media sosial yang paling banyak digunakan adalah twitter (Oktasari, 2016).

*Twitter* merupakan salah satu media sosial dan *mikroblog* yang mengizinkan setiap penggunanya untuk saling mengirim dan membaca pesan yang tidak lebih dari 280 karakter yang sering disebut dengan *tweet*. *Twitter* saat ini juga memiliki banyak pengguna dari berbagai latar belakang di mulai dari kalangan umum, *artis*, *publik figur*, hingga pejabat negara (Abdulloh dkk, 2020).

Studi kasus yang akan dibahas dalam penelitian ini yaitu tanggapan masyarakat terhadap ketua DPR Ibu Puan Maharani media yang digunakan adalah *twitter*. Namun jika dihubungkan dengan *sentiment analisis*, *twitter* itu sendiri belum memiliki fitur untuk mendeteksi cuitan yang berkaitan dengan sentimen positif dan negatif. Salah satu cara untuk mengolah data yang berupa teks yaitu *text mining* (Abdulloh dkk, 2020).

*Teks mining* bertujuan untuk mencari informasi yang penting secara terstruktur maupun tidak terstruktur yang dapat dilihat serta dimengerti oleh manusia dengan menggunakan metode yang sudah diproses oleh computer tanpa menghilangkan data penting yang sudah ada (Alhamda dkk, 2019). Salah satu contoh dalam penerapan *teks mining* yaitu klasifikasi teks. Dalam penelitian ini, klasifikasi teks digunakan untuk mengklasifikasi sebuah komentar pengguna twitter kedalam kelas Sentimen Positif dan Sentimen negatif. Dengan demikian, diperlukan suatu algoritma perhitungan yang dapat melakukan pengelompokan klasifikasi ke dalam dua kelas. Ada beberapa perhitungan algoritma yang dapat digunakan untuk pengklasifikasi seperti *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), dan *Naïve Bayes Classifier* (NBC).

Jika dikaitkan dengan penelitian ini menggunakan seleksi fitur *Information Gain*, berdasarkan penelitian Arif dkk (2020), dengan penambahan fitur *Information Gain* mampu menambah nilai akurasi sebesar 0,856.

Penelitian tentang klasifikasi sebelumnya telah dilakukan pada penelitian Candra dkk, (2020). Dalam penelitian tersebut pengujian nilai sentimen pada setiap komentar diklasifikasikan menjadi dua macam kelas. Algoritma klasifikasi yang digunakan yaitu KNN dengan menggunakan *confusion matrix* yang masing-masing *foldnya* dibentuk 4,5 dan 10 menghasilkan 95 *fold* pengujian dan menghasilkan akurasi rata-rata sebesar 58.38% dan akurasi tertinggi pada *fold* ke 6 sebesar 77%.

Dari berbagai referensi pemeriksaan yang telah diselesaikan, baik perhitungan menggunakan seleksi fitur *Information Gain* maupun *K-Nearest Neighbor* keduanya memiliki presisi akurasi yang lumayan tinggi dan sesuai untuk melakukan pengelompokan. Mengingat landasan di atas, penelitian ini akan menggunakan perhitungan yaitu *K-Nearest Neighbor* dengan seleksi fitur *Information Gain* spesifiknya untuk mengelompokkan komentar pengguna *twitter* menjadi dua kelas, yaitu sentiment positif dan sentiment negatif. Pada penelitian ini akan membandingkan tingkat hasil akurasi, dan presisi jika *K-Nearest Neighbor* menggunakan seleksi fitur *Information Gain* dan metode *K-Nearest Neighbor* tanpa seleksi fitur *Information Gain* pada saat proses klasifikasi.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah pada penelitian ini adalah:

1. Berapa akurasi dan presisi klasifikasi komentar pengguna di *twitter*.

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui tingkat akurasi dan presisi dan *K-Nearest Neighbor* dalam mengelompokkan komentar pengguna *twitter*.

## 1.4 Manfaat Penelitian

Dengan adanya penelitian ini manfaat yang diharapkan yaitu :

1. Memperoleh informasi mengenai hasil kinerja metode *K-Nearest Neighbors* tanpa seleksi fitur *Information Gain* dan metode *K-Nearest Neighbors* dengan seleksi fitur *Information Gain*.

2. Hasil penelitian ini dapat dijadikan rujukan untuk peneliti lain secara umum dengan menggunakan metode *K-Nearest Neighbors* dan seleksi fitur *Information Gain*.

### 1.5 Batasan Penelitian

Batasan masalah dari penelitian ini adalah :

1. Penelitian ini melakukan klasifikasi hanya pada komentar Bahasa Indonesia baku menurut KBBI.
2. Data yang dikumpulkan dilakukan dengan cara *crawling*. Mulai dari tanggal 3 Januari 2024 – 31 Januari 2024
3. Data yang dikumpulkan untuk dijadikan *sample* sebanyak 405 komentar *tweet*.
4. Tidak melibatkan komentar yang berisi emoji dan/atau emotikon.
5. Kata kunci yang digunakan yaitu Puan Maharani.
6. Data teks komentar diklasifikasikan menjadi dua kelas yaitu *sentimen positif* juga, *sentiment negatif*.