

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang

Pada masa pandemi sekarang ini banyaknya penggunaan media sosial yang dimana masyarakat memiliki opini yang dapat mempengaruhi bagi perkembangan masyarakat (BALYA, 2019). Media sosial sudah menjadi bagian terpenting dari percakapan diseluruh dunia, oleh karena itu masyarakat bebas mengekspresikan sebuah pendapat tentang suatu objek yang ada (Al-Khafaji and Habeeb, 2017), Maka dari itu salah satu dari media sosial yang sering digunakan oleh masyarakat salah satunya adalah Youtube. Youtube adalah sebuah situs *Web* yang memungkinkan pengguna mengunggah, menonton, dan berbagi video, youtube yang dimana sekarang setiap individu dapat mengekspresikan pendapatnya melalui bentuk video yang dapat diunggah tanpa adanya batasan waktu, Terutama berdasarkan kolom komentar yang telah disediakan oleh youtube sehingga komentar tersebut dapat dimanfaatkan untuk melakukan analisis opini publik apakah bernilai positif maupun negatif. Dalam penentuan positif atau negatif suatu opini bisa dilakukan dengan cara manual akan tetapi seiringnya banyaknya opini dari masyarakat menjadi semakin memakan banyak waktu dan usaha yang dibutuhkan untuk mengklasifikasi opini tersebut, Oleh karena itu diajukan penerapan metode pembelajaran mesin untuk melakukan klasifikasi dari sumber data opini yang banyak tersebut, Untuk melakukan hal itu menggunakan salah satu fungsi yaitu *Text Mining* yang digunakan untuk melakukan klasifikasi dokumen (Nurhuda, Widya Sihwi and Doewes, 2016).

Channel Sekretariat Presiden adalah sebuah channel youtube saluran resmi sekretariat presiden dimana isi dari video tersebut berisi tentang kegiatan presiden dan menteri yang ada di Indonesia. Channel youtube ini dibuat pada tanggal 21 januari 2019 yang rata rata statistik video sudah ditonton sebanyak 186,548,430 juta. Pada Channel Sekretariat Presiden banyak sekali video dan komentar yang beragam macam karena youtube pun membebaskan orang-orang untuk berkomentar video tersebut secara bebas dimana komentar tersebut sangat banyak sehingga membutuhkan banyak waktu untuk menilai apakah komentar tersebut bersifat positif atau negatif. Salah satu video dari channel tersebut memiliki judul “LIVE: Keterangan Pres Presiden RI Terkait Undang-Undang Cipta Kerja, Istana Bogor, 9 Oktober 2020” yang memiliki komentar sebanyak 19,134 per tanggal 25 oktober 2020.

Analisis sentimen adalah salah satu bagian dari *Text Mining*, yang dapat menganalisis pendapat seseorang, opini, evaluasi, penilaian, sikap dan emosi seseorang terhadap entitas produk, layanan, organisasi, individu, masalah, peristiwa, topik dan atribut (Liu, 2012). Analisis sentimen merupakan pengolahan kata untuk melihat atau melacak opini, evaluasi, penilaian, sikap dan emosi masyarakat tentang produk atau opini tertentu (Setiawan *et al.*, 2020). Untuk melakukan sentimen analisis dapat menggunakan komentar, komentar yang akan dianalisa akan diklasifikasi sehingga dapat terlihat apakah komentar tersebut bernilai positif atau negatif. *Text Mining* merupakan teknik yang dapat digunakan untuk melakukan sebuah penelitian klasifikasi dari sebuah dokumen yang dimana dokumen tersebut memiliki data yang berjumlah besar. *Text Mining* tersebut memiliki variasi dari data

yang dimana dari *Data Mining* tersebut akan berusaha menemukan sebuah pola dari sekumpulan data yang tekstual yang berjumlah besar (Feldman and Sanger, 2006).

Pada penelitian sebelumnya yang berjudul “Sentimen analisis pada teks bahasa indonesia menggunakan *Support Vector Machine* dan *K-Nearest Neighbors*”, menggunakan *K-Fold Cross Validation* mendapatkan akurasi KNN sebesar 60.20% dan SVM sebesar 67.90% (Lidya, Sitompul and Efendi, 2015).

Pada penelitian yang lain yang berjudul “Implementasi Analisis *Clustering* dan Sentimen *Data Twitter* pada Opini Wisata Pantai Menggunakan Metode *K-Means*, menggunakan algoritma kalsifikasi *Support Vector Machine* mendapatkan hasil akurasi sebesar 74.39% (Syarifudin and Irawan, 2018). Pada Penelitian sebelumnya dari mahasiswa Universitas Multimedia Nusantara yang berjudul “Analisis Sentimen Destinasi Wisata Labuan Bajo Pada Instagram” Menunjukkan hasil dari Metode sentimen *Lexicon-Based* menggunakan *VADER sentiment* memberikan nilai akurasi 72% (Gousander, 2020). Pada penelitian yang serupa yang berjudul “Perancangan sistem analisis sentimen komentar pelanggan menggunakan metode *Naive Bayes classifier*” mendapatkan hasil dari akurasi sistem memiliki nilai akurasi kategori 77.14% dan untuk *Precision* sentimen 99.12%, *Recall* sentimen 72.9% dan *Accuration* sentimen 75.42% (Sipayung, Maharani and Zefanya, 2016).

Berdasarkan penelitian terdahulu dari (Gousander, 2020), yang meneliti tentang “Analisis Sentimen Destinasi Wisata Labuan Bajo Pada Instagram” pada penelitian tersebut hasil dari Metode sentimen *Lexicon-Based* menggunakan *VADER sentiment* memberikan nilai akurasi 72% akan tetapi pada penelitian dari

(Sipayung, Maharani and Zefanya, 2016) yang berjudul “Perancangan sistem analisis sentimen komentar pelanggan menggunakan metode *Naive Bayes classifier*” memiliki hasil yang cukup baik nilai akurasi 77.14% , *Precision* sentimen 99.12%, *Recall* sentimen 72.9% dan *Accuration* sentimen 75.42%

Maka dari itu berdasarkan penelitian-penelitian diatas, alasan Pemilihan metode ini dikarenakan dapat memprediksi probabilitas di masa depan berdasarkan pengalaman yang sudah ada di masa lalu (Sipayung, Maharani and Zefanya, 2016). *Naive Bayes* menerapkan fungsi statistik sederhana yang berdasarkan teorema *Bayes* dengan asumsi keberadaan dari suatu fitur tertentu suatu kelas yang tidak berhubungan dengan fitur lainnya (Sari and Hayuningtyas, 2019). Pada penelitian ini juga menambahkan proses yang ada di *Preprocessing*, proses yang ditambahkan salah satunya *Normalization* yang bertujuan untuk mengubah *Teks* menjadi bentuk standar. Normalisasi *Teks* penting untuk *Teks* yang berisi seperti komentar media sosial, pesan *Teks*, dan komentar ke *Posting* blog Pada teks dengan singkatan, salah eja dan penggunaan kata kata di luar kosakata lazim (Ganesan, 2019). Pada penelitian ini juga menggunakan data ulasan komentar yang ada pada Channel Sekretariat Presiden yang berjudul “LIVE: Keterangan Pres Presiden RI Terkait Undang-Undang Cipta Kerja, Istana Bogor, 9 Oktober 2020”, Maka dari alasan tersebut dibuatlah judul skripsi “Implementasi *Naive Bayes* pada sentimen analisis komentar Youtube (Studi Kasus : Channel Sekretariat Presiden) dengan mempertimbangkan hasil akurasi penggunaan *Naive Bayes* pada analisis sentimen.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang dijelaskan sebelumnya masalah yang dapat dirumuskan adalah sebagai berikut

1. Bagaimana performa akurasi, *Precision*, *Recall*, dan *F1-score* pada *Naive Bayes Classifier* dalam melakukan analisis sentimen komentar youtube pada channel sekretariat Presiden?
2. Bagaimana mengimplementasikan algoritma *Naive Bayes* pada sentimen analisis komentar youtube (Studi Kasus : Channel Sekretariat Presiden)

## 1.3. Batasan Masalah

Batasan masalah dibuat pada implementasi ini adalah :

1. Implementasi mengklasifikasi 2 kategori yaitu komentar positif atau negatif
2. Performa akurasi diukur dengan *Confusion Matrix*
3. Analisa komen video pada channel sekretariat presiden yang berjudul “LIVE: Keterangan Pres Presiden RI Terkait Undang-Undang Cipta Kerja, Istana Bogor, 9 Oktober 2020” yang memiliki komentar sebanyak 19,134 per tanggal 25 oktober 2020.
4. Untuk *Training* model *Naive Bayes*, Kelas sentimen ditentukan oleh peneliti sebanyak 500 komentar.
5. Proses pengambilan data komentar youtube menggunakan teknik *Scraping* menggunakan bahasa pemrograman python.
6. Data yang digunakan berbahasa indonesia

7. Pada penelitian ini tidak menggunakan kelas netral , karena penentuan pada kategori netral tersebut sangat sulit untuk suatu kalimat

#### **1.4. Tujuan Penelitian**

Tujuan penelitian ini untuk melihat sentimen positif dan negatif pada komentar pengguna aplikasi youtube terdapat pada video channel sekretariat Presiden yang berjudul LIVE: Keterangan Pres Presiden RI Terkait Undang-Undang Cipta Kerja, Istana Bogor, 9 Oktober 2020” secara cepat dan otomatis dan melakukan pengukuran akurasi analisis sentimen terhadap komentar pengguna aplikasi youtube terdapat pada video tersebut menggunakan metode *Naive Bayes*.

#### **1.5. Manfaat Penelitian**

Manfaat dari penelitian ini adalah untuk dapat mengetahui kategori kalimat pada komentar tentang video yang berjudul LIVE: Keterangan Pres Presiden RI Terkait Undang-Undang Cipta Kerja, Istana Bogor, 9 Oktober 2020” apakah termaksud komentar yang bersifat positif atau negatif dengan menggunakan algoritma klasifikasi seperti *Naive Bayes* dapat efektif menilai sebuah komentar dalam bahasa indonesia.

#### **1.6. Sistematika Penulisan**

Sistematika penulisan yang digunakan dalam laporan skripsi ini adalah sebagai berikut.

## BAB 1 PENDAHULUAN

Bab ini menjelaskan latar belakang yang memiliki judul “Implementasi *Naive Bayes* pada sentimen analisis komentar Youtube (Studi Kasus : Channel Sekretariat Presiden)”, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

## BAB 2 LANDASAN TEORI

Bab ini membahas dasar teori yang akan digunakan pada penelitian ini. Teori yang digunakan pada penelitian ini yaitu *Pre-Processing Text* yang berisi *Tokenisasi, Case Folding, Filtering, Stemming, Dan Normalization, Term Frequency Inverse Document Frequency, Algoritma Naïve Bayes, K-Fold Cross Validation, dan Confusion matrix* terdiri dari akurasi, *Precision, Recall, dan F1-Score*.

## BAB 3 METODOLOGI PENELITIAN

Bab ini membahas metodologi penelitian dan perancangan aplikasi Yang terdiri dari Teknik pengumpulan Data, Flowchart dan rancangan antarmuka aplikasi.

## BAB 4 HASIL DAN DISKUSI

Bab ini membahas spesifikasi perangkat yang digunakan, implementasi sistem untuk klasifikasi komentar youtube positif atau negatif pada video channel sekretariat Presiden yang berjudul LIVE: Keterangan Pres Presiden RI Terkait Undang-Undang Cipta Kerja, Istana Bogor, 9 Oktober 2020” dan hasil implementasi

menggunakan *K-Fold Cross Validation* yang berisi hasil rata-rata nilai akurasi, *Precision*, *Recall*, dan *F1-Score*.

## BAB 5      SIMPULAN DAN SARAN

Bab ini membahas hasil simpulan dari penelitian yang telah di uji coba dan saran untuk meningkatkan dan membangun penelitian lebih lanjut.