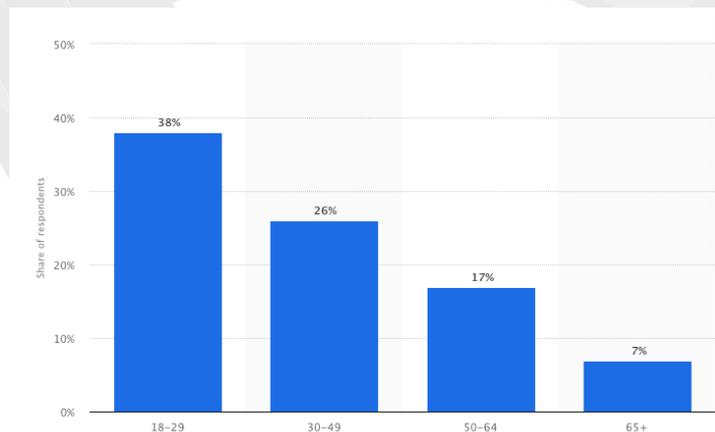


BAB I

PENDAHULUAN

1.1 Latar Belakang

Kehidupan manusia di era globalisasi ini tidak dapat dihindari dari penggunaan media sosial. Media sosial banyak digunakan karena selain dapat diakses dengan mudah, dapat digunakan sebagai wadah bagi manusia untuk mengekspresikan perasaan, ungkapan kebahagiaan atau kekecewaan, serta mendiskusikan masalah yang sedang terjadi saat ini [1]. Terdapat berbagai macam media sosial, salah satunya adalah *Twitter*.



Gambar 1.1: Persentase Pengguna *Twitter* berdasarkan Usia di Amerika Serikat tahun 2019

Sumber: *statista.com* [2]

Pengguna *Twitter* berasal dari berbagai kalangan usia. Pada Gambar 1.1, dapat dilihat bahwa pengguna *Twitter* terbanyak adalah manusia berusia 18-29 tahun. Kemampuan yang disediakan oleh *Twitter* dalam mencari berita terbaru membuat orang semakin tertarik untuk menggunakan *Twitter*. Pada media sosial ini, para pengguna diberi kebebasan untuk berekspresi, menyuarakan apa yang pengguna ingin katakan, dan berinteraksi antara satu dengan yang lain. Dengan kemampuan dari *Twitter*, maka berita yang dapat diakses akan semakin beragam dan luas. Salah satu berita yang dapat diakses pada *Twitter* adalah kegiatan politik [3].

Salah satu kegiatan politik yang sedang menjadi perbincangan saat ini adalah *United States (US) Election 2020*. *US Election* merupakan kegiatan di Amerika

Serikat dalam memilih calon presiden dan wakil presiden yang dilaksanakan pada hari Selasa setelah hari Senin pertama pada bulan November setiap empat tahun sekali [4]. Pada tahun ini, *US Election 2020* dilaksanakan pada tanggal 3 November 2020, dengan kandidat calon presiden adalah Joe Biden dan Donald Trump. Kedua kandidat ini menuai respon yang beragam, baik positif maupun negatif. Keberagaman respon yang ada dapat digunakan untuk melihat preferensi dari pengguna *Twitter* mengenai keduanya dengan menggunakan analisis sentimen [5].

Analisis sentimen merupakan penentuan pendapat atau tingkah laku manusia melalui teks yang disampaikan. Kalimat yang dianalisis diolah dengan cara memisahkan kalimat menjadi kata, kemudian dari kata tersebut akan dilihat apakah memiliki sentimen positif atau negatif. Manfaat dari analisis sentimen adalah membantu untuk melihat pendapat atau pandangan orang lain mengenai suatu kejadian [6].

Analisis sentimen dapat dilakukan dengan proses klasifikasi data berupa kalimat. Data yang berupa kalimat disebut sebagai *unstructured data* [7]. Data yang didapat akan dianalisis menggunakan *machine learning*. Terdapat dua teknik pada *machine learning*, yaitu *supervised* dan *unsupervised learning*. *Supervised learning* merupakan model yang dibentuk untuk melakukan analisis pada suatu data dengan metode KURG, sehingga metode tersebut dapat digunakan untuk melakukan pengklasifikasian data baru yang belum berlabel. *Unsupervised learning* mengolah data dengan metode untuk dapat mengklasifikasi data yang belum berlabel [8].

Teknik yang digunakan untuk melakukan analisis sentimen pada penelitian ini adalah *supervised learning*. Pemilihan penggunaan *supervised learning* ini dikarenakan pada penelitian ini dilakukan pengumpulan data yang sudah berlabel. Terdapat berbagai macam *supervised learning*, misalnya *Naive Bayes* dan *Support Vector Machine (SVM)*. Peneliti menggunakan kedua metode ini yang disimulasikan terhadap data asli yang dikumpulkan melalui *Twitter*. Kedua metode dibandingkan untuk melihat metode yang menghasilkan tingkat akurasi tertinggi untuk melakukan analisis sentimen. Dengan menggunakan data dari *Twitter* yang diambil setiap hari selama satu bulan, diharapkan hasil yang diperoleh melalui penelitian ini sesuai dengan kondisi yang sebenarnya.

1.2 Rumusan Masalah

Merujuk dari latar belakang, maka rumusan masalah yang diteliti adalah sebagai berikut.

1. Bagaimana memodelkan metode *Naive Bayes* dan *Support Vector Machine* (SVM) dalam melakukan analisis sentimen *Twitter*?
2. Bagaimana tingkat akurasi pada metode *Naive Bayes* dan SVM pada analisis sentimen *Twitter*?

1.3 Tujuan Penelitian

Tujuan yang diharapkan dapat digenapi dalam penelitian ini adalah sebagai berikut.

1. Membangun model dengan metode *Naive Bayes* dan SVM pada analisis sentimen *Twitter*.
2. Perbandingan metode *Naive Bayes* dan SVM untuk menemukan metode terbaik berdasarkan akurasi yang dapat digunakan untuk analisis sentimen *Twitter*.

1.4 Batasan Masalah

Batasan yang muncul dapat dilihat sebagai berikut.

1. Pengumpulan data pada *Twitter* menggunakan bantuan program *Rapid Miner*.
2. Pengolahan data dengan *Naive Bayes* dan SVM menggunakan program *R*.
3. Penggunaan data berasal dari *Twitter* mengenai *US Election 2020*, yang dikumpulkan dari 25 September 2020 hingga 25 Oktober 2020 sebanyak 250 *tweets* per hari.
4. Kata kunci yang digunakan pada pengumpulan data hanya berfokus pada kandidat calon presiden.

1.5 Manfaat Penelitian

Manfaat dapat dilihat dari dua segi yang dijelaskan sebagai berikut.

1.5.1 Manfaat Teoretis

Manfaat teoritis, diharapkan dengan adanya penelitian ini dapat menunjukkan perbandingan mengenai penggunaan metode *Naive Bayes* dan SVM serta metode paling optimal yang dapat digunakan untuk melakukan analisis sentimen *Twitter*.

1.5.2 Manfaat Praktis

Manfaat praktis, harapan dengan adanya penelitian ini agar dapat dijadikan sebagai acuan dan referensi bahwa pengungkapan ekspresi masyarakat di media sosial terutama *Twitter* dapat memprediksi hal yang sedang terjadi, terutama mengenai *US Election* melalui analisis sentimen.

1.6 Sistematika Penulisan

Sistematika penulisan yang terdapat di penulisan ini dapat dijabarkan sebagai berikut.

1. Dalam Bab 1 dijelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan masalah yang digunakan dalam penelitian, manfaat penelitian, sistematika penulisan, dan *timeline* penelitian.
2. Dalam Bab 2 diberikan teori-teori dasar yang digunakan dalam penulisan Skripsi mengenai metode *Naive Bayes*, metode *Support Vector Machine*(SVM), serta visualisasi model.
3. Dalam Bab 3 dijelaskan mengenai metode yang digunakan, langkah-langkah pengerjaan dari penelitian metode *Naive Bayes* dan SVM.
4. Dalam Bab 4 dijelaskan mengenai laporan dari pengolahan data dan akurasi yang diperoleh dari kedua metode beserta visualisasi data.
5. Dalam Bab 5 dipaparkan kesimpulan dan saran untuk pihak yang membutuhkan hasil penelitian ini sebagai referensi.