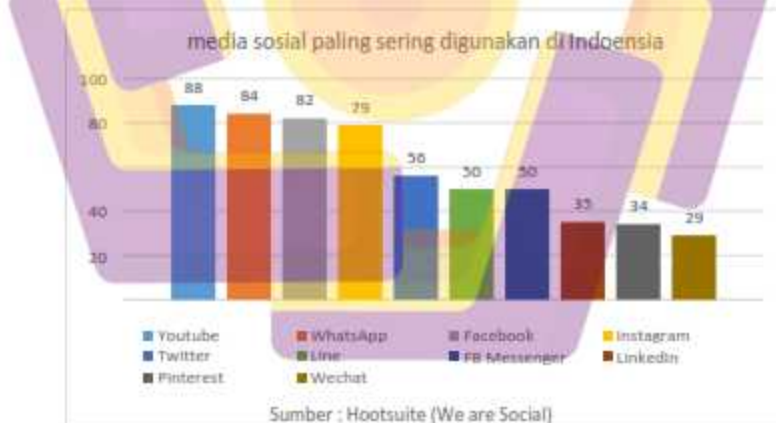


# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Pertumbuhan pengguna internet di Indonesia mengalami penambahan yang sangat tinggi tiap tahunnya, menurut laporan Asosiasi Jasa Internet Indonesia (APJII), setidaknya ada 338,2 juta koneksi seluler di Indonesia pada Januari 2020. Jumlah koneksi seluler di Indonesia meningkat sebesar 15 juta dengan kenaikan sebesar 4,6 persen, antara Januari 2019 sampai dengan Januari 2020 sudah setara dengan 124 persen dari total populasi penduduk Indonesia. Meningkatnya pengguna internet di Indonesia mempengaruhi manusia dalam melakukan interaksi sosial dimana manusia lebih menyukai interaksi sosial menggunakan media sosial.



Gambar 1.1 Daftar media sosial yang sering digunakan di Indonesia

Perkembangan internet kini memengaruhi keberadaan sosial media yang semakin banyak, aplikasi sosial media bahkan kini menjadi suatu yang tidak dapat lepas dari kehidupan masyarakat. Terutama untuk kalangan remaja (generasi milenial).

Terkait dengan populernya media sosial, Loddixon Gonzales dengan temanya ingin menciptakan *platform* media sosial yang nyaman dan pas dihati masyarakat Indonesia, sehingga terciptalah sebuah aplikasi bernama 'sestyc'. Sestyc untuk pertama kalinya dirilis *playstore* pada 1 april 2019 dan tanggal 1 april 2020 dirilis di *App Store*.

Banyaknya algoritma yang digunakan untuk melakukan sentimen analisis, peneliti memilih menggunakan algoritma *Support Vector Machine (SVM)*, *Logistic Regression*, *Naive bayes*.

*Support Vector Machine* merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning (supervised learning) yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan oleh Vladimir Vapnik padatahun 1992. Klasifikasi dilakukan dengan garis pembatas (*hyperlane*) yang memisahkan antara kelas opini positif dan opini negatif. Secara intuitif, suatu garis pembatas yang baik adalah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas, karena pada umumnya semakin besar margin, semakin rendah error generalisasi dari pemilah. Margin adalah jarak dari suatu titik vektor di suatu kelas terhadap *hyperplane*[1].

*Logistic Regression* adalah sebuah algoritma klasifikasi untuk mencari hubungan antara variabel tak bebas dikotomus (berskala dengan dua kategori) atau polikotomus (berskala dengan lebih dari dua kategori) dengan variabel bebas bersifat kontinu atau kategorik[2].

*Naive Bayes Classifier* merupakan metode klasifikasi dengan menggunakan metode probabilitas dan statistik. Algoritma ini bekerja dengan melakukan perhitungan probabilitas dengan menjumlahkan kombinasi dan frekuensi dari data uji. Dengan metode teorema *bayes*, dengan mengasumsikan ketergantungan variabel satu dengan variabel lainya dengan klasifikasi[3].

Pada penelitian ini akan dilakukan Sentimen analisis menggunakan 3 algoritma untuk mengetahui performa terbaik dari algoritma tersebut dengan menggunakan sentimen pengguna *sestyc*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana cara menggunakan Algoritma klasifikasi untuk melakukan analisis sentimen terhadap pengguna *sestyc*?
2. Berapa cara menentukan algoritma terbaik untuk melakukan klasifikasi sentiment terhadap review *sestyc*?

### 1.3 Batasan Masalah

1. Data yang digunakan merupakan review yang di ambil dari website google play store.
2. Metode analisis klasifikasi sentiment menggunakan tiga algoritma, yaitu : *support vector machine, naïve bayes, dan logistic regression*

### 1.4 Tujuan Penelitian

Tujuan penelitian adalah :

1. Dapat menerapkan analisis sentimen pada pengguna aplikasi Sestyc menjadi kelas positif dan negatif dengan menggunakan metode *Support Vector Machine, Naive Bayes, dan Logistic Regression*.
2. Menganalisis dan menemukan algoritma terbaik untuk melakukan klasifikasi sentimen pada review sestyc.
3. Dapat mengetahui berapa akurasi yang didapatkan dari penelitian.

### 1.5 Manfaat Penelitian

Adapun manfaat penelitian sebagai berikut :

1. Bagaimana cara menggunakan Algoritma klasifikasi untuk melakukan analisis sentimen terhadap pengguna sestyc?
2. Mengetahui tanggapan pengguna dan sentimen terhadap aplikasi Sestyc
3. Menjadi acuan untuk perbandingan performa algoritma klasifikasi *Support vector machine, naïve bayes dan logistic regression*.
4. Dapat dijadikan referensi bagi peneliti lain.

## 1.6 Sistematika Penulisan

Untuk mengetahui uraian singkat yang memuat gambaran singkat secara keseluruhan isi masing-masing bab, maka dibuat sistematika sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini berisi tentang uraian latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

### **BAB II LANDASAN..TEORI**

Bab ini berisi tentang teori-teori yang digunakan serta konsep dasar yang menjadi landasan dari penelitian ini yang berasal dari berbagai sumber seperti jurnal dan proceeding.

### **BAB III METODE PENELITIAN**

Bab ini berisi tentang analisis dan perancangan sistem dari metode yang digunakan

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini membahas tentang proses implementasi yang dilakukan secara menyeluruh serta hasil yang didapatkan dari penelitian.

### **BAB V PENUTUP**

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk memperbaiki kekurangan yang ada pada penelitian.