

# PERBANDINGAN ALGORITMA NAÏVE BAIYES CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM MENGANALISIS SENTIMEN TERHADAP PELAYANAN GOJEK

Annisa Nur Salsabila, Munirul ula\* dan Yesy Afrillia

*Program Studi Teknik Informatika, Universitas Malikussaleh, Aceh, Indonesia.*

*\*Email: munirulula@unimal.ac.id*

## Abstrak

Di era digital ini, masyarakat banyak meminati transportasi online yang berbasis aplikasi. Salah satu transportasi online yang sangat populer di Indonesia adalah Gojek. Bertambahnya minat pengguna layanan Gojek menjadikannya perbincangan di media sosial. Masyarakat memberikan tanggapan berupa keluhan, kritikan, saran maupun pujian terhadap pelayanan Gojek. Menganalisa opini masyarakat ini secara manual membutuhkan waktu yang lama. Maka dari itu, diperlukannya analisis sentimen untuk mengekstrak opini – opini tersebut. Dalam menganalisis sentimen diperlukannya sebuah teknik yaitu text mining. Text mining merupakan suatu proses pengambilan intisari dari dokumen teks sehingga didapatkan hasil yang berguna untuk tujuan tertentu. Dalam penelitian ini dilakukan perbandingan algoritma Naïve Baiyes dan Support Vector Machine (SVM) untuk menganalisis sentimen terhadap pelayanan gojek. Penelitian ini bertujuan untuk mengimplementasikan algoritma naïve baiyes classifier dan support vector machine dalam mengklasifikasi sentimen pengguna terhadap pelayanan Gojek Indonesia serta menganalisa tingkat keakuratan data tertinggi antara algoritma naïve bayes dan support vector machine. Penelitian ini menggunakan 1175 tweet yang terdiri dari 608 tweet label positif, 317 tweet label negatif, dan 249 tweet label netral. Kemudian menggunakan komposisi data sebesar 70% data training dan 30% data testing, yaitu sebanyak 822 tweet yang digunakan sebagai data training dan 353 tweet sebagai data testing. Berdasarkan hasil penelitian diperoleh performa SVM lebih bagus dibandingkan dengan Naive Baiyes dengan tingkat akurasi 58,35%, presisi sebesar 75,3%, recall sebesar 43,67%, dan f1-score sebesar 41,67%. Sedangkan Naïve baiyes menghasilkan akurasi 53,54%, presisi sebesar 84%, recall sebesar 36%, dan f1-score sebesar 29%.

**Kata kunci:** *Gojek, Analisis Sentimen, Text Mining, Naïve Baiyes, Support Vector Machine.*

## Pendahuluan

Di era digital ini, masyarakat banyak meminati transportasi online yang berbasis aplikasi. Salah satu transportasi online yang sangat populer di Indonesia adalah Gojek dengan berbagai fitur yang memberikan kemudahan dan menghemat waktu. Disamping itu, transportasi online belum sepenuhnya sempurna, sehingga diperlukannya review pengguna untuk meningkatkan kepuasan pengguna terhadap pelayanan Gojek. Hal ini dapat diketahui dari kualitas layanan, kecepatan, biaya, serta keamanan.

Bertambahnya minat pengguna layanan Gojek menjadikannya perbincangan di media sosial. Masyarakat memberikan tanggapan berupa keluhan, kritikan, saran maupun pujian terhadap pelayanan Gojek. Opini dari masyarakat ini dijadikan evaluasi bagi perusahaan – perusahaan transportasi online dalam meningkatkan

kualitas pelayanannya. Salah satu caranya dengan menganalisa opini – opini tersebut sehingga mengetahui seberapa puas pengguna terhadap pelayanan yang ditawarkan. Menganalisa opini masyarakat ini secara manual membutuhkan waktu yang lama. Maka dari itu, diperlukannya analisis sentimen untuk mengekstrak opini – opini tersebut sehingga mendapatkan informasi sentimen yang terkandung dalam sebuah opini teks, baik itu sentimen positif, negatif maupun netral.

Analisis sentimen adalah bidang studi yang menganalisis pendapat seseorang, opini, evaluasi, penilaian, sikap dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik dan atribut. Dengan analisis sentimen kita dapat menentukan opini, emosi dan sikap yang dicerminkan melalui teks, biasanya diklasifikasikan menjadi opini negatif dan positif [1].

Dalam menganalisis sentimen diperlukannya sebuah teknik yaitu text mining. Text mining merupakan suatu proses pengambilan intisari dari dokumen teks sehingga didapatkan hasil yang berguna untuk tujuan tertentu. Ada beberapa algoritma yang dapat digunakan dalam analisis sentimen diantaranya yang paling populer adalah Naïve Bayes dan Support Vector Machine.

Naïve Bayes merupakan algoritma yang sangat stabil, lebih baik, dan memiliki kinerja yang lebih cepat sehingga membuatnya sangat populer. algoritma ini digunakan untuk berbagai bidang klasifikasi. Naïve Bayes berkinerja cukup konsisten dan bagus dalam pengaturan eksperimental yang berbeda. Hal ini sederhana untuk implementasi dan independensi memungkinkan parameter menjadi diperkirakan pada kumpulan data yang berbeda [2]. Naïve Bayes dapat digunakan untuk menarik kesimpulan dalam kasus-kasus dengan banyak sumber pengukuran yang tidak dapat ditangani oleh metode lain seperti model hierarkis yang kompleks [3].

Support Vector Machine (SVM) merupakan salah satu metode pengklasifikasian yang memberikan hasil terbaik [4]. SVM memiliki kemampuan yang lebih baik dalam klasifikasi daripada metode lain. Dengan adanya kemampuan generalisasi, SVM mampu menghasilkan akurasi yang tinggi dan tingkat kesalahan yang relatif kecil. Pada perkembangannya, SVM telah berhasil digunakan untuk menyelesaikan permasalahan dalam berbagai bidang, diantaranya adalah klasifikasi pada microarray data [5].

Gojek merupakan perusahaan yang didirikan oleh anak bangsa yang bernama Nadiem Makarim bersama temannya Michaelangelo Moran pada bulan Maret 2014. Gojek merupakan salah satu jasa penyedia transportasi online yang mampu mempertahankan pangsa pasarnya pada kondisi persaingan transportasi online yang semakin ketat [6].

Penelitian ini bertujuan untuk mengimplementasikan algoritma naïve bayes classifier dan support vector machine dalam mengklasifikasi sentimen pengguna terhadap pelayanan Gojek Indonesia serta menganalisa tingkat keakuratan data tertinggi antara algoritma naïve bayes dan support vector machine.

## **Tinjauan Pustaka**

Preprocessing merupakan tahapan awal yang berfungsi untuk mentransformasikan dokumen ke dalam bentuk representasi yang lain. Preprocessing dapat berguna untuk menghilangkan noise, menyeragamkan bentuk kata dan mengurangi volume kosakata Tujuannya adalah untuk mempercepat proses terhadap dokumen, mempermudah proses pencarian query ke dalam dokumen, dan mempermudah

proses pengurutan dokumen yang terambil Preprocessing ini terdiri dari beberapa tahap, yaitu:

### **Cleaning**

*Cleaning* ialah adalah proses yang digunakan untuk melakukan pembersihan text dari karakter-karakter selain huruf, tanda baca dan tag yang tidak digunakan nantinya di dalam proses information retrieval.

#### a. *Case Folding*

*Case Folding* merupakan proses dalam melakukan penyeragaman kata kedalam data tweet. Proses ini juga melakukan perubahan huruf kapital menjadi huruf kecil termasuk juga penghapusan setiap karakter tanda baca.

#### b. *Tokenization*

Tokenisasi adalah adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Dimana deretan kata dalam kalimat, paragraf, atau halaman dipisah menjadi token atau potongan kata tunggal.

#### c. *Filtering*

*Filtering* adalah proses penghapusan *stopwords*. *Stopword* adalah teks yang sering muncul di dalam dokumen tapi tidak mengandung informasi yang signifikan. Dengan dihapusnya *stopwords*, ukuran kosakata menjadi berkurang sehingga hanya kata-kata penting yang terdapat dalam dokumen.

#### d. *Stemming*

Stemming adalah merupakan proses pemotongan imbuhan atau pengembalian kata berimbuhan menjadi kata dasar. Tujuan utama stemming adalah untuk mengubah susunan kata seperti kata benda, kata sifat, kata kerja, kata keterangan, dan sebagainya ke dalam bentuk kata dasar

### **Naïve Baiyes Classifier**

*Naïve Bayes Classifier* adalah pengklasifikasi statistik, karena klasifikasi didasarkan pada perhitungan probabilitas, dan pengetahuan domain juga diekspresikan secara probabilistik [12]. *Naive bayes* merupakan salah satu algoritma yang dapat digunakan untuk memprediksi keanggotaan dari suatu class berdasarkan teorema *bayes* yang mampu bekerja seperti *decision tree* dan *neural network*. *Naive bayes* melakukan klasifikasi secara efektif dengan mengoptimalkan pengawasan perkiraan dalam probabilitas yang akurat, bahkan ketika asumsi dilanggar performa *classifier* ini tetap lebih baik dibandingkan *classifier* yang lain. [13]. Berikut bentuk persamaan teorema bayes :

$$P(X|H) = \frac{P(H|X) \times P(X)}{P(H)} \quad (1)$$

Dimana:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X
- P(H) : Probabilitas hipotesis H
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Dalam proses membangun sistem pengklasifikasi menggunakan algoritma Naïve Baiyes terdapat dua tahapan yang dilakukan sebagai berikut :

### 1. Proses Training

Pada tahap ini dilakukan pelatihan menggunakan data latih berupa komentar yang telah diproses dan yang sudah ditentukan label (kelas) untuk membangun sebuah model. Berikut langkah-langkahnya :

- a. Menghitung probabilitas kata (TF-IDF)
- b. Menghitung probabilitas kategori dari data latih

$$P(v_j) = \frac{\text{jumlah dokumen dengan label } v_j}{\text{keseluruhan dokumen}} \quad (2)$$

- c.  $P(a_i|v_j) = \frac{pk+1}{n+term}$  Menghitung probabilitas untuk setiap kelas data latih (3)

Dimana,

- $P(a_i|v_j)$  = probabilitas term
- $pk + 1$  = jumlah kemunculan kata  $a_i$  terhadap kategori  $v_j$ , penambahan 1 dilakukan agar hasil tidak 0
- $n$  = semua jumlah kata yang muncul pada kategori  $v_j$
- $term$  = jumlah kata keseluruhan

### 2. Proses Testing

Tahap ini dilakukan untuk mengetahui keakuratan model yang dibangun pada tahap training. Pada tahap ini digunakan data yang disebut data uji untuk memprediksi label (kelas) dari data model. Berikut langkah-langkahnya :

- a. Menentukan  $V_{map}$  untuk setiap kelas data uji

$$V_{map} = P(v_j) \times P(a_i|v_j) \quad (4)$$

- b. Menentukan  $V_{map}$  maksimum

- c. Diperoleh kelas prediksi

Jika dari perhitungan data uji tersebut  $V_{map}$  Positif  $>$   $V_{map}$  Negatif maka termasuk ke dalam kategori kelas positif.

## Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan machine learning (supervised learning) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training. Klasifikasi dilakukan dengan mencari hyperplane atau garis pembatas (decision boundary) yang memisahkan antara suatu kelas dengan kelas lain, yang dalam kasus ini garis tersebut berperan memisahkan tweet bersentimen positif (berlabel +1) dengan tweet bersentimen negatif (berlabel -1). SVM melakukan pencarian nilai hyperplane dengan menggunakan support vector dan nilai margin [14].

### Klasifikasi Non-Linear SVM

Pada umumnya masalah dalam dunia nyata jarang bersifat linear separable. Kebanyakan bersifat non linear. Untuk itu, SVM dimodifikasi dengan memasukkan fungsi kernel. Kernel trick adalah salah satu cara yang dapat membantu untuk memudahkan dalam melakukan klasifikasi dalam bentuk non linear, cukup dengan hanya memahami atau mengetahui fungsi kernel apa yang cocok digunakan dalam kasus linear yang dihadapi. Berikut beberapa fungsi kernel yang digunakan dalam SVM sebagai berikut:

- a. Linear

$$K(x_i, x) = x_i^T x \quad (5)$$

b. Polinomial

$$K(x_i, x) = (\gamma(x_i^T x) + r)^p \quad (6)$$

c. Radial Basis Function (RBF)

$$K(x_i, x) = \exp(-\gamma\|x_i - x_j\|^2) \quad (7)$$

d. Sigmoid

$$K(x_i, x) = \tanh(\gamma(x_i^T x) + r) \quad (8)$$

Pada penelitian ini data yang digunakan berbentuk teks yang bersifat real world problem sehingga proses klasifikasi tidak dapat diasumsikan secara linear, oleh karena itu akan digunakan kernel sebagai parameter yang dapat melakukan klasifikasi terhadap data non linear. Kernel yang digunakan adalah Kernel RBF dengan parameter  $c$  dan  $\gamma$ . [15] merekomendasikan fungsi kernel RBF untuk digunakan karena kemampuannya dalam mengklasifikasikan data berdimensi tinggi dan juga memiliki parameter yang lebih sedikit daripada kernel polinomial.

### SMO-SVM

Bentuk primal problem SVM yang sulit diselesaikan bisa diubah kedalam bentuk dual yang mengandung nilai  $\alpha$ . Banyak algoritma dikembangkan guna menemukan nilai  $\alpha$  akan tetapi algoritma-algoritma yang ada memerlukan waktu yang lama apalagi jika jumlah data training yang dipakai berukuran besar karena algoritma tersebut menggunakan metode *numerical quadratic programming* sebagai *inner loop* [16]. Maka dari itu, untuk mengatasi masalah tersebut digunakan metode *Sequential Minimal Optimization* (SMO). Berikut langkah-langkah penyelesaian training :

1. Menghitung perhitungan kernel terlebih dahulu

2. Menghitung matriks dengan rumus berikut :

$$D_{ij} = Y_i Y_j (K(x_i x_j) + \lambda^2) \quad (9)$$

Keterangan :

$D_{ij}$  = Elemen matriks ke ij

$Y_i$  = Kelas data i (label)

$Y_j$  = Kelas data j (label)

$\lambda^2$  = Batas teoritis yang diturunkan

3. Menghitung nilai error

$$E_i = \sum_{j=1}^n a_j D_{ij} \quad (10)$$

Keterangan :  $E_i$  = Nilai error data ke i

4. Menghitung delta  $a_i$

$$\delta a_i = \min\{\max[\gamma(1 - E_i) - a_i], C - a_i\} \quad (11)$$

Keterangan :

$\delta a_i$  = Delta  $\alpha$  ke i

$\gamma$  = Gamma

$C$  = Complexity

5. Menghitung  $a_i$  baru

$$a_i \text{ baru} = a_i + \delta a_i \quad (12)$$

6. Menghitung nilai  $w \cdot x^+$  dan  $w \cdot x^-$  untuk mendapatkan nilai bias

$$w \cdot x^+ = a_i Y_i K(w \cdot x^+) \text{ dan } w \cdot x^- = a_i Y_i K(w \cdot x^-) \quad (13)$$

$$b = \frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (14)$$

Keterangan :

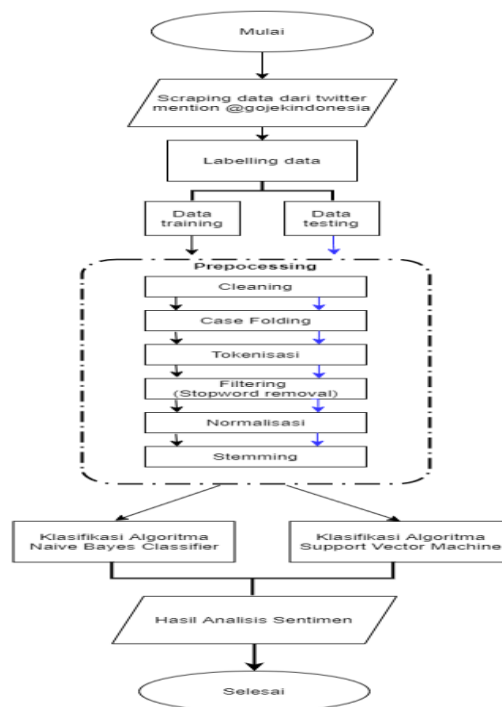
$w \cdot x^+$  = Nilai kernel data x dengan data x kelas positif yang memiliki nilai  $\alpha$

- Tertinggi  
 $w \cdot x^-$  = Nilai kernel data x dengan data x kelas negatif yang memiliki nilai  $\alpha$   
 Tertinggi
7. Menghitung nilai keputusan  

$$f(x) = \sum_{i=1}^m \text{sign}(a_i Y_i K(x, x_i) + b) \quad (15)$$
  8. Ulangi langkah 3-5 hingga  $a_i$  baru konvergen (tidak ada perubahan signifikan).

### Metodologi Penelitian

Data diperoleh dari kumpulan tweet mention @gojekindonesia menggunakan API twitter dan library tweepy. Namun karena keterbatasan akses API twitter yang gratis sehingga scraping data yang diperoleh kurang akurat sehingga penulis juga menggunakan cara manual dalam mengumpulkan data yaitu dengan menggunakan fitur search pada twitter untuk melakukan pencarian data tweet dengan menggunakan lebih dari satu keyword, sehingga diperoleh 1176 tweet. Kemudian dilakukan pemberian label secara manual. Maka didapatkan hasil dengan 608 tweet label positif, 317 tweet label negatif, dan 249 tweet label netral. Kemudian data tersebut dibagi menjadi data latih sebanyak 70% dan data uji sebanyak 30% yang dipilih secara acak.



Gambar 1 Skema Sistem

## Hasil dan Pembahasan

Data komentar yang telah selesai melalui proses preprocessing akan dilakukan proses TF-IDF. Setiap kata atau term yang sudah melalui tahap preprocessing dilakukan pemberian nilai atau bobot. Pada TF-IDF perhitungan bobot sebuah term dalam sebuah komentar dilakukan pengkalian nilai Term Frequency dengan Inverse Document Frequency.

### Perhitungan Manual Algoritma Naïve Baiyes

#### a. Proses Training

Tabel 1. Contoh Kasus Data Training

No.	Tweet	Sentimen
1.	sayang banget gojek banget promo	Positif
2.	kemarin order makan gojek map akurat terima order makan rugi layan main buruk mahal	Negatif
3.	halo akun kena shadow banned kah tidak bisa pakai promo	Netral

Dari data tabel 2 dibuat sebuah model probabilitas dengan mengacu pada persamaan 3 sebagai berikut :

$$P(a_{sayang}|V_{positif}) = \frac{1 + 1}{5 + 29} = \frac{2}{34}$$

$$P(a_{sayang}|V_{negatif}) = \frac{0 + 1}{14 + 29} = \frac{1}{43}$$

$$P(a_{sayang}|V_{netral}) = \frac{0 + 1}{10 + 29} = \frac{1}{39}$$

Jika dibuat menjadi sebuah tabel, maka probabilitas setiap kata pada data training seperti tabel 3 berikut :

Tabel 2. Perhitungan Probabilitas Data Training

Sentimen	P(vj)	P(ai Vj)			
		sayang	banget	gojek	promo
Positif	1/3	2/36	3/36	2/36	2/36
Negatif	1/3	1/46	1/46	1/46	1/46
Netral	1/3	1/52	1/52	1/52	1/52

Tabel 3. Perhitungan Probabilitas Data Training (Lanjutan)

Senti men	P( vj)	P(ai Vj)													
		kem arin	order	ma kan	goj ek	ma p	aku rat	teri ma	rugi	lay an	mai n	bur uk	ma hal	pul a	

Pos	1/3	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36
Neg	1/3	2/46	3/46	3/46	2/46	2/46	2/46	2/46	2/46	2/46	2/46	2/46	2/46	2/46
Net	1/3	1/42	1/42	1/42	1/42	1/42	1/42	1/42	1/42	1/42	1/42	1/42	1/42	1/42

Tabel 4. Perhitungan Probabilitas Data Training (Lanjutan)

Sentimen	P(vj)	P(ai Vj)												
		Halo	akun	ken a	shado w	banne d	kah	tida k	bisa	paka i	prom o			
Pos	1/3	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36	1/36
Neg	1/3	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46
Net	1/3	2/42	2/42	2/42	2/42	2/42	2/42	2/42	2/42	2/42	2/42	2/42	2/42	2/42

Hasil perhitungan probabilitas tersebut digunakan sebagai model probabilistik yang selanjutnya digunakan sebagai data acuan untuk menentukan data testing.

#### b. Proses Testing

Tabel 5. Contoh Kasus Data Testing

No.	Tweet	Sentimen
1.	jujurly murah aman gosend banget promo pakai toped kayak ya allah belanja mulu saya	?

Proses testing ini dihitung probabilitasnya dan dicari probabilitas tertinggi menggunakan persamaan 4 sebagai berikut :

- $$P(\text{"Positif"}|\text{"T4"}) = P(\text{"Positif"}) \times P(\text{"jujurly"}|\text{"positif"}) + P(\text{"murah"}|\text{"positif"}) + P(\text{"aman"}|\text{"positif"}) + P(\text{"gosend"}|\text{"positif"}) + P(\text{"banget"}|\text{"positif"}) + P(\text{"promo"}|\text{"positif"}) + P(\text{"pakai"}|\text{"positif"}) + P(\text{"toped"}|\text{"positif"}) + P(\text{"kayak"}|\text{"positif"}) + P(\text{"ya"}|\text{"positif"}) + P(\text{"allah"}|\text{"positif"}) + P(\text{"belanja"}|\text{"positif"}) + P(\text{"mulu"}|\text{"positif"}) + P(\text{"saya"}|\text{"positif"})$$

$$= \frac{1}{3} \times (\frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34} + \frac{1}{34})$$

$$= 0,13725$$

- $$P(\text{"Negatif"}|\text{"T4"}) = P(\text{"Negatif"}) \times P(\text{"jujurly"}|\text{"negatif"}) + P(\text{"murah"}|\text{"negatif"}) + P(\text{"aman"}|\text{"negatif"}) + P(\text{"gosend"}|\text{"negatif"}) + P(\text{"banget"}|\text{"negatif"}) + P(\text{"promonya"}|\text{"negatif"}) + P(\text{"pakai"}|\text{"negatif"}) + P(\text{"toped"}|\text{"negatif"}) + P(\text{"kayak"}|\text{"negatif"}) + P(\text{"ya"}|\text{"negatif"}) + P(\text{"allah"}|\text{"negatif"}) + P(\text{"belanja"}|\text{"negatif"}) + P(\text{"mulu"}|\text{"negatif"}) + P(\text{"saya"}|\text{"negatif"})$$

$$= \frac{1}{3} \times (\frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43} + \frac{1}{43})$$

$$= 0,10853$$



- $$\begin{aligned}
 P(\text{"Netral"}|\text{T4}) &= P(\text{"Netral"}) \times P(\text{"jujurly"}|\text{"netral"}) + P(\text{"murah"}|\text{"netral"}) + \\
 &P(\text{"aman"}|\text{"netral"}) + P(\text{"gosend"}|\text{"netral"}) + P(\text{"banget"}|\text{"netral"}) + \\
 &P(\text{"promonya"}|\text{"netral"}) + P(\text{"pakai"}|\text{"netral"}) + P(\text{"toped"}|\text{"netral"}) + \\
 &P(\text{"kayak"}|\text{"netral"}) + P(\text{"ya"}|\text{"netral"}) + P(\text{"allah"}|\text{"netral"}) + P(\text{"belanja"}|\text{"netral"}) + \\
 &P(\text{"mulu"}|\text{"netral"}) + P(\text{"saya"}|\text{"netral"}) \\
 &= 1/3 \times (1/39 + 1/39 + 1/39 + 1/39 + 1/39 + 2/39 + 2/39 + 1/39 + 1/39 + 1/39 + \\
 &1/39 + 1/39 + 1/39 + 1/39) \\
 &= 0,13675
 \end{aligned}$$

Dengan melakukan perhitungan probabilitas teks terhadap sentimen dengan menggunakan algoritma Naïve Baiyes maka dapat diketahui bahwa teks tweet T4 termasuk dalam kategori sentimen positif.

### Perhitungan Manual Algoritma Support Vector Machine (SVM)

#### 1. Proses Training

- Melakukan Kernelisasi dengan Kernel RBF dengan nilai  $\gamma = 0,5$ . Sehingga membentuk matriks sebagai berikut :

$$K(x_i, x_j) = \begin{bmatrix} 1 & 0.2689 & 0.3108 \\ 0.2689 & 1 & 0.1082 \\ 0.3108 & 0.1082 & 1 \end{bmatrix}$$

- Perhitungan Hessian matriks

$$x_1x_1 = (1)(1)(1 + 0.5^2) = 1.25$$

$$x_1x_2 = (1)(-1)(0.2689 + 0.5^2) = -0.5189$$

$$x_1x_3 = (1)(0)(0.3108 + 0.5^2) = 0$$

$$x_2x_3 = (-1)(0)(0.2689 + 0.5^2) = 0$$

Matriksnya :

$$D_{ij} = \begin{bmatrix} 1.25 & -0.5189 & 0 \\ -0.5189 & 1.25 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Mencari nilai error

$$x_1 = (1.25 - 0.5189 + 0) \times 0.5 = 0.3656$$

$$x_2 = (-0.5189 + 1.25 + 0) \times 0.5 = 0.3656$$

$$x_3 = (0 + 0 + 0) \times 0.5 = 0$$

Matriksnya :

$$E_i = \begin{bmatrix} 0.3656 \\ 0.3656 \\ 0 \end{bmatrix}$$

- Menghitung delta alpha

$$x_1 = \text{Min}(\text{Max}(0.5(1 - 0.3656), -0.5), 1 - 0.5)$$

$$= \text{Min}(0.3172, 0.5)$$

$$= 0.3172$$

Matriksnya :

$$\delta\alpha_i = \begin{bmatrix} 0.3172 \\ 0.3172 \\ 0 \end{bmatrix}$$

- Menghitung nilai alpha baru

$$x_1 = 0.5 + 0.3172 = 5.31723$$

Sehingga keadaan konvergen diulangi sampai iterasi ke 3, maka matriksnya :

$$\alpha_i = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

f. Mencari  $w \cdot x^+$  dan  $w \cdot x^-$  untuk mencari nilai bias

- Nilai positif terletak pada data ke-1

$$\begin{aligned} w \cdot x^+ &= ((1 \times 1 \times 1) + ((-1) \times 1 \times 0.2689) + (0 \times 1 \times 0.3108)) \\ &= (1 - 0.2689 + 0) \\ &= 0.7311 \end{aligned}$$

- Nilai negatif terletak pada data ke-2

$$\begin{aligned} w \cdot x^- &= ((1 \times 1 \times 0.2689) + ((-1) \times 1 \times 1) + (0 \times 1 \times 0.1082)) \\ &= (0.2689 - 1 + 0) \\ &= -0.7311 \end{aligned}$$

g. Menghitung nilai bias

$$b = -\frac{1}{2}(0.7311 + (-0.7311)) = 0$$

2. Proses Testing

Misalkan data uji berupa komentar yang telah diproccessing sehingga menghasilkan vektor sebagai berikut :

$$x_3 = \begin{bmatrix} 0,0,0,0,0.176,0, \dots, 0.477, 0.477, 0.477, 0.477, 0.477, 0.477, \\ 0.477, 0.477, 0.477 \end{bmatrix}$$

- Melakukan perhitungan kernel RBF, perhitungan dilakukan sampai data ke 3,

Tabel 6. Hasil Dot Product Data Uji

$K(x, x_1)$	$K(x, x_2)$	$K(x, x_3)$
0,3108	0,10823	1

- Untuk mengetahui kelas data uji maka dihitung fungsi keputusan  $sign()$

$$D(x) = sign((1 \times 1 \times 0.3108) + ((-1) \times 1 \times 0.1082) + (0 \times 1 \times 1))$$

$$(x) = sign(0.2026)$$

$$(x) = 1 \text{ (positif)}$$

Dari hasil perhitungan diatas komentar yang telah dilakukan pengujian merupakan kelas positif.

## Evaluasi Klasifikasi

Setelah proses klasifikasi dilakukan diatas maka tahap selanjutnya yaitu tahap evaluasi klasifikasi. Dimana pada tahap ini dilakukan pengujian dengan menggunakan Confusion Matrix.

1. Confusion Matrix Naïve Baiyes

Tabel 7. Confusion Matrix Naïve Baiyes

Aktual	Prediksi		
	Positif	Negatif	Netral
Aktual Positif	179	0	0
Aktual Negatif	85	9	0
Aktual Netral	79	0	1

Dari tabel diatas dapat dihitung akurasi, presisi, recal, dan f1-score nya. Berikut perhitungannya :

$$Akurasi = \frac{179 + 9 + 1}{179 + 85 + 9 + 79 + 1} \times 100\% = \frac{189}{353} \times 100\% = 53,54\%$$

$$\begin{aligned} Presisi &= \frac{Presisi(pos) + Presisi(neg) + Presisi(net)}{3} \times 100\% \\ &= \frac{0,52 + 1 + 1}{3} \times 100\% = 84\% \end{aligned}$$

$$\begin{aligned} Recall &= \frac{Recall(pos) + Recall(neg) + Recall(net)}{3} \times 100\% \\ &= \frac{1 + 0,09 + 0,0125}{3} \times 100\% = 36\% \end{aligned}$$

$$\begin{aligned} F1 - Score &= \frac{F1 - Score(pos) + F1 - Score(neg) + F1 - Score(net)}{3} \times 100\% \\ &= \frac{0,68 + 0,165 + 0,025}{3} \times 100\% = 29\% \end{aligned}$$

## 2. Confusion Matriks SVM

Tabel 8. Confusion Matrix SVM

Aktual	Prediksi		
	Positif	Negatif	Netral
Aktual Positif	176	3	0
Aktual Negatif	71	22	1
Aktual Netral	70	2	8

Dari tabel diatas dapat dihitung akurasi, presisi, recal, dan f1-score nya. Berikut perhitungannya :

$$Akurasi = \frac{176 + 22 + 8}{179 + 94 + 80} \times 100\% = \frac{206}{353} \times 100\% = 58,35\%$$

$$\begin{aligned} Presisi &= \frac{Presisi(pos) + Presisi(neg) + Presisi(net)}{3} \times 100\% \\ &= \frac{0,56 + 0,81 + 0,89}{3} \times 100\% = 75,3\% \end{aligned}$$

$$\begin{aligned} Recall &= \frac{Recall(pos) + Recall(neg) + Recall(net)}{3} \times 100\% \\ &= \frac{0,98 + 0,23 + 0,10}{3} \times 100\% = 43,67\% \end{aligned}$$

$$F1 - Score = \frac{F1 - Score(pos) + F1 - Score(neg) + F1 - Score(net)}{3} \times 100\%$$

$$= \frac{0,71 + 0,36 + 0,18}{3} \times 100\% = 41,67\%$$

## Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma Naïve Baiyes dan Support Vector Machine dapat melakukan analisis sentimen terhadap pelayanan gojek. Penelitian ini menggunakan 1175 data, dimana 821 data training dan 353 data testing. Hasil yang diperoleh dapat dinyatakan bahwa performa SVM lebih bagus dibandingkan dengan Naive Baiyes dengan tingkat akurasi 58,35%, presisi sebesar 75,3%, recall sebesar 43,67%, dan f1-score sebesar 41,67%. Sedangkan Naïve baiyes menghasilkan akurasi 53,54%, presisi sebesar 84%, recall sebesar 36%, dan f1-score sebesar 29%. Dari hasil dan pembahasan diatas juga disimpulkan bahwa di media twitter memberikan respon positif yang lebih besar daripada respon negatif terhadap pelayanan gojek yang artinya kepuasan pengguna gojek lebih banyak daripada keluhannya.

## Daftar Pustaka

- [1] B. Liu, *Sentiment Analysis and Opinion Mining Mining. Synthesis Lectures on Human Language Technologies. [Draft]*, vol. 5, no. 1. 2012.
- [2] T. Subramaniam, H. A. Jalab, and A. Y. Taqa, "Overview of textual anti-spam filtering techniques," *Int. J. Phys. Sci.*, vol. 5, no. 12, pp. 1869–1882, 2010.
- [3] J. Simangunsong, "Universitas Sumatera Utara Skripsi," *Anal. Kesadahan Total dan Alkalinitas pada Air Bersih Sumur Bor dengan Metod. Titrim. di PT Sucofindo Drh. Provinsi Sumatera Utara*, no. Cmc, pp. 44–48, 2019.
- [4] Y. Chen, C. Lu, C. Huang, and P. R. a Pproaches, "Anti-Spam Filter Based on Naïve Bayes ," pp. 1–5, 2009.
- [5] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000, doi: 10.1093/bioinformatics/16.10.906.
- [6] F. Anggraini and A. Budiarti, "Pengaruh Harga, Promosi, dan Kualitas Pelayanan Terhadap Loyalitas Pelanggan Dimediasi Kepuasan Pelanggan Pada Konsumen Gojek," *J. Pendidik. Ekon.*, vol. 8, no. 3, pp. 86–94, 2020, doi: 10.26740/jupe.v8n3.p86-94.
- [7] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [8] S. Gusriani, K. D. K. Wardhani, and M. I. Zul, "Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)," *4th Appl. Bus. Eng. Conf.*, vol. 1, no. 1, pp. 1–7, 2017, [Online]. Available: [https://www.researchgate.net/profile/Muhammad-Ihsan-Zul/publication/311107797\\_Analisis\\_Sentimen\\_Terhadap\\_Toko\\_Online\\_di\\_Sosial\\_Media\\_Menggunakan\\_Metode\\_Klasifikasi\\_Naive\\_Bayes\\_Studi\\_Kasus\\_Facebook\\_Page\\_BerryBenka/links/583e787b08ae8e63e6158698/Analisis-Se](https://www.researchgate.net/profile/Muhammad-Ihsan-Zul/publication/311107797_Analisis_Sentimen_Terhadap_Toko_Online_di_Sosial_Media_Menggunakan_Metode_Klasifikasi_Naive_Bayes_Studi_Kasus_Facebook_Page_BerryBenka/links/583e787b08ae8e63e6158698/Analisis-Se).
- [9] F. Syadid, "Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia

- 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf- Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network,” *Skripsi Univ. Islam Negeri Syarif Hidayatullah Jakarta*, p. 72, 2019.
- [10] I. Fera Fanesya, Randy Cahya Wihandika, “Deteksi Emosi pada Twitter Menggunakan Metode Naive Bayes dan Kombinasi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, p. 3, 2019.
- [11] S. S. Salim and J. Mayary, “Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor,” *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.
- [12] T. Mitsa, *Temporal data mining*. 2010.
- [13] Fadhil, “Berformalin Berbasis Android Berdasarkan Image Mata Menggunakan Metode Naive Bayes Classifier,” *Skripsi Univ. Islam Negeri Maulana Malik Ibrahim Malang*, 2016.
- [14] A. Novantirani, M. K. Sabariah, and V. Effendy, “Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine,” *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.
- [15] S. Ding and L. Chen, “Intelligent Optimization Methods for High-Dimensional Data Classification for Support Vector Machines,” *Intell. Inf. Manag.*, vol. 02, no. 06, pp. 354–364, 2010, doi: 10.4236/iim.2010.26043.
- [16] J. Platt, “Sequential minimal optimization: A fast algorithm for training support vector machines. Advances in Kernel Methods-Support Vector learning,” *Cambridge, MA MIT Press*, no. July 1998, pp. 185–208, 1999.