

# KLASIFIKASI KONTEN INSTAGRAM BERDASARKAN KOMENTAR MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Daniar Nur Amin<sup>1</sup>, Ednawati Rainarli<sup>2</sup>

<sup>1,2</sup> Teknik Informatika – Universitas Komputer Indonesia

Jl. Dipatiukur 112-114 Bandung

E-mail: daniarnuramin@email.unikom.ac.id<sup>1</sup>, ednawati.rainarli@email.unikom.ac.id<sup>2</sup>

## ABSTRAK

Instagram merupakan salah satu aplikasi media sosial yang populer digunakan oleh banyak orang di seluruh dunia salah satunya di Indonesia. Berdasarkan fungsinya aplikasi ini digunakan untuk berbagi foto maupun video pada sesama pengguna. Dengan semakin banyaknya pengguna maka postingan yang ada dalam instagram perlu dikelompokkan berdasarkan kategorinya. Penelitian ini menggunakan komentar untuk mengelompokkan postingan gambar pada Instagram. Metode *Support Vector Machine* dipilih untuk mengklasifikasikan teks pada komentar tersebut. Pada tahapan *preprocessing* ada yang berbeda dari penelitian yang sudah dilakukan yaitu pada tahapan normalisasi bahasa karena pada tahapan ini diubah kata yang tidak baku menjadi kata yang baku sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Pengujian dilakukan menggunakan metode *Support Vector Machine* dengan kernel linear, pengujian dilakukan dengan data latih sebanyak 200 data dan untuk data uji sebanyak 100 data, dan hasil akurasi yang didapatkan dari pengujian yaitu sebesar 96%. Berdasarkan hasil tersebut maka disimpulkan bahwa *Support Vector Machine* dapat digunakan untuk mengklasifikasikan konten Instagram berdasarkan komentar.

**Kata kunci:** Instagram, Postingan, Konten Instagram, Klasifikasi, *Support Vector Machine*

## 1. PENDAHULUAN

Instagram merupakan salah satu aplikasi media sosial yang populer digunakan oleh banyak orang di seluruh dunia salah satunya di Indonesia. Berdasarkan fungsinya aplikasi ini digunakan untuk berbagi foto maupun video pada sesama pengguna. Dengan semakin banyaknya pengguna maka postingan yang ada dalam instagram perlu dikelompokkan berdasarkan kategorinya [1]. Dari setiap postingan di instagram terdapat banyak komentar yang dilakukan oleh sesama pengguna. Komentar yang ada pada postingan ini berpotensi digunakan untuk mengelompokkan gambar yang dikirim ke dalam kategori-kategori tersebut. Oleh karena itu dibutuhkan suatu cara untuk mengelompokkan konten instagram berupa postingan gambar berdasarkan komentar.

Penelitian terhadap klasifikasi atau pengelompokan berupa data teks berbahasa Indonesia pernah dilakukan sebelumnya oleh Dio Ariandi, mengenai klasifikasi berita Indonesia menggunakan metode *Naïve Bayes Classification* dan *Support Vector Machine*. Hasilnya diperoleh dengan menggunakan metode *Support Vector Machine* 88,1% lebih baik 5,9% dari NBC dengan akurasi 82.2% [2]. Peneliti lainya mengklasifikasikan berita *online* menggunakan metode *Support Vector Machine* dan *K-Nearest Neighbor* oleh Siti Nur Aisyah. Hasil akurasi metode *Support Vector Machine* didapatkan sebesar 93.2% lebih baik 33.2% dibandingkan dengan *K-Nearest Neighbor* dengan akurasi 60% [3]. Penelitian yang sudah dilakukan menggunakan metode *Support Vector Machine* memiliki nilai akurasi yang baik, maka dari itu pada penelitian ini akan menggunakan metode *Support Vector Machine* untuk membantu mengklasifikasikan data yang sudah ditentukan.

Metode *Support Vector Machine* adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi, penelitian ini akan mengimplementasikan metode *Support Vector Machine* untuk mengklasifikasikan konten instagram berdasarkan komentar pada suatu postingan.

### 1.1 Klasifikasi

Klasifikasi adalah proses pengelompokan objek yang memiliki karakteristik atau ciri yang sama ke dalam beberapa teks[4]. Konten Instagram berdasarkan komentar merupakan proses untuk mengklasifikasi atau mengelompokkan data berupa komentar kedalam kategori tertentu, yang dimana ini berkaitan dengan *text mining*. Pengertian dari *text mining* itu sendiri yaitu proses penemuan akan informasi atau trend baru yang sebelumnya tidak terungkap dengan memproses dan menganalisa data dalam jumlah besar. Dalam menganalisa sebagian atau keseluruhan unstructured text, *text mining* mencoba untuk mengasosiasikan satu bagian text dengan yang lainnya berdasarkan aturanaturan tertentu. Hasil yang di harapkan adalah informasi baru atau “insight” yang tidak terungkap jelas sebelumnya [5][6].

## 1.2 Preprocessing

*Preprocessing* merupakan tahap pertama yang dilakukan adalah proses pelatihan. Dalam proses pelatihan data latih akan melalui beberapa proses utama yaitu proses *preprocessing*, pembobotan dan pelatihan *Support Vector Machine*. Proses *preprocessing* itu sendiri terdiri dari enam proses yaitu *Case Folding*, *Cleansing*, *Filtering*, *Tokenizing*, *Normalisasi Bahasa* dan *Stopwords Removal*.

### 1.2.1 Case Folding

*Case Folding* adalah proses menyeragamkan bentuk kata-kata pada komentar menjadi huruf kecil (*lowercase*) atau huruf kapital (*uppercase*). Pada penelitian ini diseragamkan menjadi bentuk *lowercase* [6].

### 1.2.2 Cleansing

*Cleansing* merupakan proses membersihkan kata-kata yang tidak diperlukan untuk mengurangi noise. Kata-kata yang dihilangkan yaitu URL, *hashtag* (#) dan *username* (@)[7].

### 1.2.3 Filtering

*Filtering* merupakan proses menghapuskan karakter-karakter selain huruf a sampai z. pada tahap ini karakter selain huruf dihilangkan [8].

### 1.2.4 Tokenizing

*Tokenizing* merupakan tahap pemotongan kalimat berdasarkan tiap kata yang menyusunnya. Pada tahap ini dilakukan penguraian deskripsi yang semula berupa kalimat menjadi kata tunggal [9].

### 1.2.5 Normalisasi Bahasa

*Normalisasi Bahasa* dilakukan terhadap kata yang tidak baku, tahapan ini bertujuan untuk mengembalikan bentuk penulisan dari masing-masing kata yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Proses ini dilakukan dengan mencocokkan setiap kata pada dokumen data latih maupun data uji dengan kata yang ada pada *List* kata yang sudah ditentukan[7][10].

### 1.2.6 Stopwords Removal

*Stopword Removal* yaitu proses membuang kata-kata yang terdapat pada daftar *stopword*. Kata-kata yang berada dalam daftar *stopword* berupa kata ganti orang, kata ganti penghubung dan penunjuk [3][11].

## 1.3 Pembobotan TF-IDF

*Term Frequency–Inverse Document Frequency* digunakan untuk menentukan nilai munculnya setiap kata dari dokumen-dokumen yang ada, perhitungan ini menggunakan rumus untuk menentukan tingkat kebenaran dari sebuah kata yang ada didalam dokumen[12].

Pembobotan didapatkan dari banyaknya kata atau term yang muncul dalam setiap dokumen yang biasa disebut *term frequency*(*tf*). *tf-idf* berhasil digunakan dalam penyaringan di berbagai bidang, termasuk *text summarization* dan klasifikasi [12].

Bobot suatu istilah semakin besar jika istilah tersebut sering muncul dalam suatu dokumen dan semakin kecil jika istilah tersebut muncul dalam banyak dokumen [12].

Nilai *idf* sebuah *term*(kata) dapat dihitung menggunakan persamaan (1) berikut:

$$idf = \log(N/df) \quad (1)$$

Untuk menghitung bobot (*W*) masing-masing dokumen terhadap system *term* (kata) dapat menggunakan persamaan (2) berikut.

$$Wt = tf \cdot idf \quad (2)$$

## 1.4 Support Vector Machine

*Support Vector Machine* (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linear dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistic. SVM adalah salah satu teknik yang relatif baru dibandingkan dengan teknik lain, tetapi memiliki performansi yang lebih baik di berbagai bidang aplikasi seperti *bioinformatics*, pengenalan tulisan tangan, klasifikasi teks, prediksi dan lain sebagainya [13][14]. Proses pembelajaran pada SVM bertujuan untuk mendapatkan hipotesis berupa bidang pemisah (*hyperplane*) terbaik. *Hyperplane* terbaik tidak hanya dapat memisahkan data tetapi juga memiliki *margin* (jarak) yang paling besar. Data yang berada pada *hyperplane* disebut *support vector*.

Data pada ruang input (*input space*) berdimensi *d* dinotasikan dengan  $x_i = \in \mathbb{R}^d$  sedangkan label kelas dinotasikan dengan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, \dots, n$ . Dimana *n* adalah banyaknya data. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara linear bidang pembatas [15], maka persamaan bidang pembatasnya didefinisikan pada persamaan (3) berikut:

$$w \cdot x_i + b = 0 \quad (3)$$

Data  $x_i$  yang terbagi ke dalam dua kelas, yang termasuk kelas -1 (sampel negatif) didefinisikan sebagai vektor yang memenuhi pertidaksamaan (4) berikut:

$$w \cdot x_i + b < 0 \text{ untuk } y_i = -1 \quad (4)$$

Sedangkan yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan (5) berikut:

$$w \cdot x_i + b > 0 \text{ untuk } y_i = +1 \quad (5)$$

Dimana:

$x_i$  = data masukan

$y_i$  = label dari setiap data

*w* = nilai dari bidang normal

$b$  = posisi bidang relatif terhadap pusat koordinat

Parameter  $w$  dan  $b$  adalah parameter yang akan dicari nilainya. Bila label data  $y_i = -1$ , maka pembatas menjadi persamaan (6) berikut:

$$w \cdot x_i + b \leq -1 \quad (6)$$

Bila label data  $y_i = +1$ , maka pembatas menjadi persamaan (7) berikut:

$$w \cdot x_i + b \geq +1 \quad (7)$$

Margin terbesar dapat dicari dengan cara memaksimalkan jarak antar bidang pembatas kedua kelas dan titik terdekatnya, yaitu  $2/|w|$ . Hal ini dirumuskan sebagai permasalahan *quadratic programming* (QP) *problem* yaitu mencari titik minimal persamaan (8) dengan memperhatikan persamaan (9) berikut:

$$\min \tau(w) = \frac{1}{2} \|w\|^2 \quad (8)$$

$$y_i(w \cdot x_i + b) - 1 \geq 0, (i = 1, \dots, n) \quad (9)$$

Permasalahan ini dapat dipecahkan dengan berbagi teknik komputasi. Lebih mudah diselesaikan dengan mengubah persamaan (8) ke dalam fungsi *Lagrangian* pada persamaan (10), dan menyederhanakannya menjadi persamaan (11) berikut:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T x_i + b - 1) \quad (10)$$

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T x_i + b) + \sum_{i=1}^n a_i \quad (11)$$

Dimana  $a_i$  adalah *lagrange multiplier* yang bernilai nol atau positif ( $a_i \geq 0$ ). Nilai optimal dari persamaan (12) dapat dihitung dengan meminimalkan  $L$  terhadap  $w$ ,  $b$  dan  $a$ . Dapat dilihat pada persamaan (13) sampai (14) berikut:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n a_i y_i x_i = 0 \quad (12)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n a_i y_i = 0 \quad (13)$$

$$\frac{\partial L}{\partial a} = \sum_{i=1}^n a_i y_i (w^T x_i + b) - \sum_{i=1}^n a_i = 0 \quad (14)$$

Dimana  $n$  adalah jumlah data yang menjadi *support vector*. Karena *Lagrange Multiplier* ( $\alpha$ ) tidak diketahui nilainya, persamaan di atas tidak dapat diselesaikan secara langsung untuk mendapatkan  $w$  dan  $b$ . Untuk menyelesaikan masalah tersebut, modifikasilah Persamaan (10) diatas menjadi kasus memaksimalkan dengan syarat optimal untuk dualitas menggunakan konstrain KKT (*Karush-Kuhn-Tucker*) sebagai berikut:

$$\text{Syarat 1: } a_i [y_i (w \cdot x_i + b) - 1] = 0 \quad (15)$$

$$\text{Syarat 2: } a_i \geq 0, i = 1, 2, \dots, n \quad (16)$$

Maka masalah *Lagrange* untuk klasifikasi dapat dinyatakan pada persamaan (17) berikut:

$$\min L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T x_i + b) - \sum_{i=1}^n a_i \quad (17)$$

Model persamaan (15) diatas merupakan model primal *Lagrange*. Sedangkan dengan memaksimalkan  $L$  terhadap  $a_i$ , persamannya menjadi persamaan (18) berikut:

$$L_D = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1, j=1}^n a_i a_j y_i y_j x_i x_j \quad (18)$$

$$\text{Syarat 1: } \sum_{i=1}^n a_i y_i = 0 \quad (19)$$

$$\text{Syarat 2: } a_i \geq 0, i = 1, 2, \dots, n \quad (20)$$

Dengan  $x_i x_j$  merupakan *dot-product* dua data dalam data latih. *Hyperplane* (batas keputusan atau pemisah) didapatkan dengan persamaan (21) berikut:

$$\left( \sum_{i=1}^n a_i y_i x_i \cdot z \right) + b = 0 \quad (21)$$

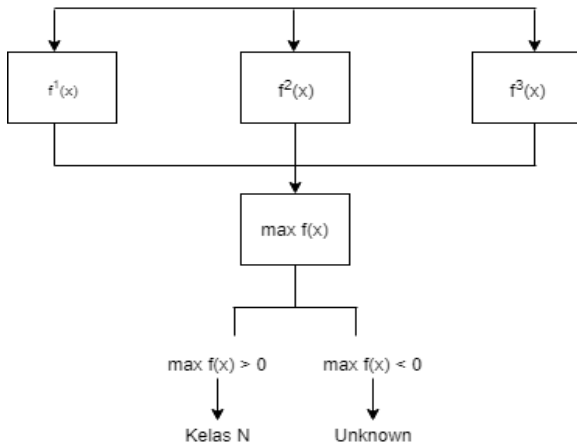
$n$  adalah banyaknya data,  $x_i$  merupakan *support vector*,  $z$  merupakan data yang akan di uji dan ditentukan kelasnya, dan  $x_i z$  merupakan *inner-product* antara  $x_i$  dan  $z$ . Untuk nilai  $b$  didapatkan dari Persamaan (15) pada *support vector*. Karena  $a_i$  dihitung dengan metode numerik dan mempunyai eror numerik, nilai yang dihitung untuk  $b$  bisa jadi tidak sama. Untuk mendapatkan  $b$ . Persamaan (15) dapat disederhanakan menjadi persamaan (22) berikut:

$$b_i = 1 - y_i (w \cdot x_i) \quad (22)$$

SVM biasanya digunakan untuk persoalan klasifikasi dengan banyaknya kelas hanya 2, kemudian coba dikembangkan untuk mengklasifikasikan lebih dari 2 kelas. Salah satu metode pendekatan yang digunakan untuk mengklasifikasikan lebih dari 2 kelas yaitu *One Against All* (OAA). Metode OAA untuk kasus klasifikasi  $k$ -kelas, menemukan  $k$  *hyperplane* dimana  $k$  adalah banyak kelas dan  $\rho$  adalah *hyperplane*. Dalam metode ini  $\rho^{(\ell)}$  diujikan dengan semua data dari kelas  $\ell$  dengan label  $+1$ , dan semua data dari kelas lain dengan label  $-1$  [8]. Berikut ini merupakan ilustrasi untuk persoalan klasifikasi dengan tiga buah jumlah kelas, digunakan tiga buah SVM biner pada Tabel 1 Dan penggunaannya pada pengklasifikasian data baru pada Gambar 1 [9].

**Tabel 1 Contoh Metode One Against All**

$y_i = 1$	$y_i = -1$	Hipotesis
Kelas 1	Bukan kelas 1	$f^1(x) = (w^1)x + b^1$
Kelas 2	Bukan kelas 2	$f^2(x) = (w^2)x + b^2$
Kelas 3	Bukan kelas 3	$f^3(x) = (w^3)x + b^3$



**Gambar 1. Metode Klasifikasi SVM One Against All**

Konsep metode pendekatan OAA yaitu dimisalkan pada tiga kelas, kelas 1,2 dan 3. Bila kelas 1 akan di ujikan, Maka kelas 1 akan diberi label +1 serta kelas 2 dan 3 diberi label -1. Bila kelas 2 akan di ujikan, Maka kelas 2 akan diberi label +1 serta kelas 1 dan 3 diberi label -1. Serta ini sama untuk kelas 3. Maka akan didapat *hyperplane* untuk masing-masing kelas di atas. Kemudian kelas dari suatu data baru  $x$  ditentukan berdasarkan nilai terbesar dari *hyperplane* [8]:

$$\text{kelas } x = \arg \max_{\ell=1 \dots k} ((w^{(\ell)})^T \cdot \phi(x) + b^{(\ell)}) \quad (23)$$

### 1.5 Pengujian Kinerja Klasifikasi

*Matrix confusion* merupakan tabel yang mencatat hasil kerja klasifikasi. Pada tabel berikut merupakan contoh *matrix confusion* yang melakukan klasifikasi masalah tiga kelas [16]. Setiap set  $F_{ij}$  dalam matriks menyatakan jumlah record/data dari kelas  $i$  yang hasil prediksinya masuk ke kelas  $j$ . Misalnya sel  $F_{11}$  adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1. Dan  $F_{12}$  adalah jumlah data dalam kelas 1 yang dipetakan ke kelas 2.

**Tabel 2 Matrix Confussion**

$F_{ij}$		Kelas Prediksi (j)		
		Kelas 1	Kelas 2	Kelas 3
Kelas Asli (i)	Kelas 1	$F_{11}$	$F_{12}$	$F_{13}$
	Kelas 2	$F_{21}$	$F_{22}$	$F_{23}$
	Kelas 3	$F_{31}$	$F_{32}$	$F_{33}$

Ketepatan klasifikasi dapat dilihat dari akurasi klasifikasi. Akurasi klasifikasi menunjukkan performansi model klasifikasi secara keseluruhan, dimana semakin tinggi akurasi klasifikasi hal ini semakin baik performansi model klasifikasi.

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

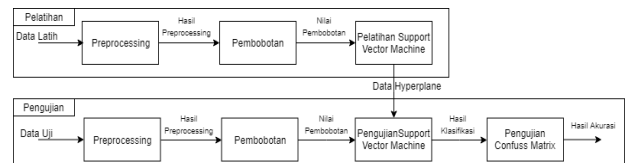
## 2. ISI PENELITIAN

### 2.1 Analisis Masalah

Berdasarkan dari perumusan masalah yang sudah didapat yaitu banyaknya pengguna jejaring sosial instagram menyebabkan banyak juga yang membuat postingan dan komentar. Sebagai pengguna tidak akan mengetahui postingan tersebut masuk kedalam kategori apa, jika hanya melihat dari komentar yang ada dan harus melihat gambar atau redaksi dari postingan, sedangkan jika ingin mengetahui postingan tersebut tentang apa hanya dengan melihat komentar-komentar yang ada. Solusi yang diberikan adalah dengan digunakannya *Machine Learning* dalam mengklasifikasikan komentar yang ada didalam instagram untuk dikategorikan sesuai dengan gambar yang ada dalam postingan.

### 2.2 Analisis Proses

Analisis proses merupakan tahapan untuk menganalisis suatu cara atau menganalisis metode-metode yang digunakan. Tahapan yang digunakan untuk mengetahui kalimat tersebut masuk kedalam konten apa sesuai dengan yang sudah ditentukan, pada penelitian ini dibagi menjadi dua tahap, yaitu tahap pelatihan dan pengujian. Berikut gambaran tahapan proses yang akan dilakukan:



**Gambar 2 Analisis Proses**

### 2.3 Analisis Data Masukan

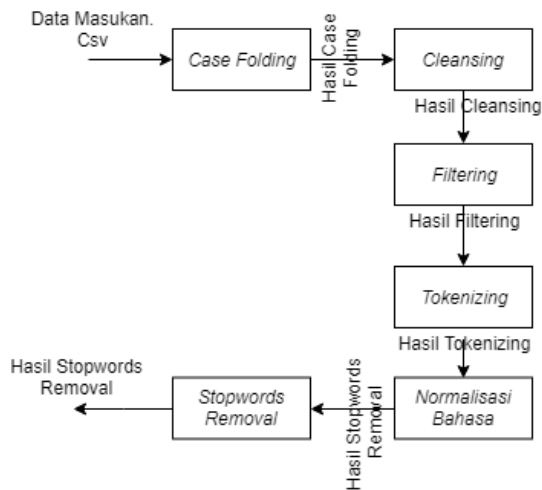
Data masukan yang digunakan adalah komentar di instagram dari postingan yang berkaitan dengan konten yang sudah ditentukan. Data komentar tersebut didapat dengan memanfaatkan teknik scraping. Data yang digunakan yaitu dari kategori yang sudah ditentukan dengan format csv.

Tabel 3 Contoh Data Masukan

Kategori	Komentar
Game	Yap dan game nya seru ?? sangat di sayangkan versi PC nya gak ada ???? @adefirsyah namanya aja ps4 exclusive ?? @adefirsyah le Sony : eh lupa gua sabar aja gan gua juga males rilis di PC ?? Gw beli nih game wktu winter sale hrge 150K. Smpe skrang blum gw mainin krna blom di download?? Inet lemot Bangsad, gua baru beli ps4 karena gue pengen maen horizon zero dawn.. Masa sekuel terbaru nya ada di ps5?? Auto misquen gue bgsd Kasih gratis donk, biar banyak yg maen termasuk aku @zairulrizal48 bukannya kalo peli ps4 langsung dapet gratis yak? Blm bisa disebut game terbaik klo ga/blm rilis di PC ?????? @ari_arifin1 God of War (ps exclusive) game terbaik 2018

2.4 Analisis Preprocessing

Analisis *Preprocessing* merupakan tahap pertama yang dilakukan adalah proses pelatihan. Dalam proses pelatihan data latih akan melalui beberapa proses utama yaitu proses *preprocessing*, pembobotan dan pelatihan *Support Vector Machine*. Proses *preprocessing* itu sendiri terdiri dari enam proses yaitu *Case Folding*, *Cleansing*, *Filtering*, *Tokenizing*, *Normalisasi Bahasa* dan *Stopwords Removal*.



Gambar 4 Analisis Preprocessing

Setelah melakukan tahap *Case Folding*, *Cleansing*, *Filtering*, *Tokenizing*, *Normalisasi Bahasa* dan *Stopwords Removal* maka hasil dari proses preprocessing terdapat pada tabel 4.

Tabel 4 Hasil Preprocessing

Kategori	Hasil Preprocessing		
Game	game seru sayangkan versi pc ps exclusive le sony lupa sabar males rilis pc game waktu winter	sale harga bermain download inet lemot beli ps main horizon zero dawn sekuel terbaru ps auto miskin	kasih gratis main peli ps gratis game terbaik rilis pc god of war ps exclusive game terbaik

2.5 Analisis Pembobotan TF-IDF

Proses awal dilakukan perhitungan *tf*(*term frequency*) pada setiap dokumenya, selanjutnya mencari nilai *df*, karena *df* merupakan banyaknya dokumen yang didalamnya muncul suatu *term*. Setelah mendapatkan nilai *df*, maka dilakukan perhitungan *idf* dengan persamaan (1).

$$idf\ t = \log (N/df) \tag{1}$$

Diambil contoh pada kata “game”. Didapatkan banyak dokumen (N) = 6, dan *df* = 2. Maka, dihitung seperti berikut.

$$idf\ t = \log(6/2) = 0,477$$

Selanjutnya untuk mendapatkan bobot term maka dilakukan perhitungan *tf* dan *idf* dengan persamaan (2)

$$Wt = tf\ d\ t * idf\ t \tag{2}$$

Diperoleh *tf* = 4, dan *idf* = 0,368. Maka perhitungannya sebagai berikut.

$$Wd\ t = 4 * 0,368 = 1,472$$

Sehingga kata “game” memiliki bobot 1,472

2.6 Analisis Pelatihan Support Vector Machine

Matriks dari setiap elemennya merupakan hasil  $x_i x_j$  yang akan berkorelasi dengan  $a_i a_j$  dalam persamaan . Dengan menggunakan matriks K sebagai *dot-product*  $x_i x_j$  dalam persamaan dualitas Lagrange multiplier, didapatkan:

Maksimalkan:

$$L_D\ max = a_1 + a_2 + a_3 + a_4 + a_5 + a_6 - \frac{1}{2} (53,486a_1^2 + 3,386a_1a_2 - 0,406a_1a_4 - 0,178a_1a_5 - 0,085a_1a_6 + 37,082a_2^2 - 0,294a_2a_3 - 0,043a_2a_5 - 0,085a_2a_6 + 64,723a_3^2 + 4,735a_3a_4 + 128,369a_4^2 + 0,677a_4a_5 + 0,271a_4a_6 + 48,275a_5^2 + 10,056a_5a_6 + 57,419a_6^2)$$

$$\text{Syarat 1: } a_1 + a_2 - a_3 - a_4 - a_5 - a_6 = 0$$

$$\text{Syarat 2: } a_1, a_2, a_3, a_4, a_5, a_6 \geq 0$$

Dalam fungsi tujuan, suku kedua sudah dikalikan dengan  $y_i y_j$ . Persamaan tersebut memenuhi standar Quadratic Programming sehingga bisa dibantu penyelesaiannya dengan solver komersial untuk Quadratic Programming (QP). Dengan bantuan perangkat lunak, didapatkan hasil sebagai berikut:

$$\begin{aligned} a_1 &= 0.190 \\ a_2 &= 0.028 \\ a_3 &= 0.130 \\ a_4 &= 0.006 \\ a_5 &= 0.015 \\ a_6 &= 0.013 \\ b &= -61.337 \end{aligned}$$

Hasil ini menunjukkan bahwa semua data training adalah support vector dikarenakan nilai  $a > 0$ . Sementara nilai  $b$  didapatkan dari proses pelatihan yang dilakukan. Setelah menemukan semua  $a$  dan  $b$ , maka model SVM dapat digunakan untuk model prediksi dengan menggunakan:

$$f(x) = w^T \cdot x + b$$

Dimana  $w^T \cdot x + b = a^T y K(x_i \cdot x_{Uji}) + b$ .

Maka persamaan  $f(x)$  adalah sebagai berikut:

$$w^T \cdot x + b = a^T y K(x_i \cdot x_{Uji}) + b.$$

$$f(x) = \begin{pmatrix} 0.190 \\ 0.028 \\ 0.130 \\ 0.006 \\ 0.015 \\ 0.013 \end{pmatrix}^T \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} K(x_1, x_{Uji}) \\ K(x_2, x_{Uji}) \\ K(x_3, x_{Uji}) \\ K(x_4, x_{Uji}) \\ K(x_5, x_{Uji}) \\ K(x_6, x_{Uji}) \end{pmatrix} - 61.337$$

## 2.7 Analisis Pengujian Support Vector Machine

Setelah ditemukan nilai data hasil dari perhitungan  $x_{Uji} x_1^T$  sampai dengan  $x_{Uji} x_6^T$ . Nilai data uji tersebut disubsitusikan ke dalam persamaan (23) berikut:

$$\begin{aligned} kelas x = \arg \max_{k=1..3} & ([w^1]^T \cdot \varphi(x) \\ & + b^1, [w^2]^T \cdot \varphi(x) \\ & + b^2, [w^3]^T \cdot \varphi(x) + b^3) \end{aligned}$$

Nilai  $w^1$  dan  $b^1$  didapatkan dari hasil pelatihan yang telah dilakukan sebelumnya dimana angka 1 menunjukkan indeks kelas 1 yaitu Game, angka 2 menunjukkan indeks kelas 2 yaitu Makanan, angka 3 menunjukkan indeks kelas 3 yaitu Olahraga.

$$kelas x = \arg \max_{k=1} \left( \begin{pmatrix} 0.190 \\ 0.028 \\ 0.130 \\ 0.006 \\ 0.015 \\ 0.013 \end{pmatrix}^T \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 5,688 \\ 5,588 \\ 0 \\ 0,296 \\ 0,085 \\ 0,171 \end{pmatrix} - 61,337 \right)$$

$$\begin{pmatrix} 0.171 \\ 0.171 \\ -0.171 \\ -0.171 \\ -0.171 \\ -0.171 \end{pmatrix}^T \begin{pmatrix} 5,688 \\ 5,588 \\ 0 \\ 0,296 \\ 0,085 \\ 0,171 \end{pmatrix} - 61,337 = 1,834 - 61,337 = -59,503$$

$$kelas x = \arg \max_{k=2} \left( \begin{pmatrix} 0,007 \\ 0,011 \\ 0,022 \\ 0,011 \\ 0,008 \\ 0,006 \end{pmatrix}^T \begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 5,688 \\ 5,588 \\ 0 \\ 0,296 \\ 0,085 \\ 0,171 \end{pmatrix} + 24,355 \right)$$

$$\begin{pmatrix} -0,001 \\ -0,001 \\ 0,001 \\ 0,001 \\ -0,001 \\ -0,001 \end{pmatrix}^T \begin{pmatrix} 5,688 \\ 5,588 \\ 0 \\ 0,296 \\ 0,085 \\ 0,171 \end{pmatrix} + 24,355 = -0,011 + 24,355 = 24,344$$

$$kelas x = \arg \max_{k=3} \left( \begin{pmatrix} 0,011 \\ 0,017 \\ 0,010 \\ 0,005 \\ 0,023 \\ 0,019 \end{pmatrix}^T \begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 5,688 \\ 5,588 \\ 0 \\ 0,296 \\ 0,085 \\ 0,171 \end{pmatrix} + 36,144 \right)$$

$$\begin{pmatrix} 0,001 \\ 0,001 \\ 0,001 \\ 0,001 \\ -0,001 \\ -0,001 \end{pmatrix}^T \begin{pmatrix} 5,688 \\ 5,588 \\ 0 \\ 0,296 \\ 0,085 \\ 0,171 \end{pmatrix} - 61,337 = 0,011 + 36,144 = 36,155$$

$$kelas x = \arg \max_{k=1..3} \begin{pmatrix} -59,503, & 24,344, \\ & 36,144 \end{pmatrix}$$

$$kelas x = 36,144$$

Nilai *hyperplane* terbesar adalah 36,144 dimana nilai *hyperplane* tersebut merupakan nilai *hyperplane* kelas 3 Berarti data  $P_{Uji}$  termasuk konten instagram dengan kategori Olahraga.

## 2.8 Pengujian dan Akurasi

Pengujian dilakukan untuk mengetahui tingkat akurasi klasifikasi konten Instagram berdasarkan komentar menggunakan *Support Vector Machine*. Pengujian akurasi merupakan tahap yang memiliki tujuan untuk mengetahui nilai akurasi dari penggunaan metode *Support Vector Machine*. Caranya yaitu dengan menghitung jumlah data uji yang diprediksi secara benar dengan metode ini. Dan untuk mengukur kinerja dengan menggunakan matriks konfusi. Pengujian dilakukan menjadi dua tahapan, yang pertama dengan menguji data latih dan data uji yang sama dengan jumlah data sebanyak 200. Dan yang kedua dengan menggunakan data latih sebanyak 200 dan data uji sebanyak 100.

- Hasil Pengujian Akurasi Pertama  
Berikut hasil pengujian pertama dengan menggunakan data latih dan data uji yang sama dengan banyaknya data 200, nilai akurasi yang didapatkan bisa dilihat Tabel 5 berikut:

**Tabel 5 Hasil Pengujian Akurasi Pertama**

Kondisi	Linear	RBF		
		$\gamma=1$	$\gamma=2$	$\gamma=3$
SVM	100%	100%	100%	100%
		Polynomial		
		n=1	n=2	n=3
		100%	100%	100%

Hasil pengujian tersebut menunjukkan bahwa metode Support Vector Machine dengan semua kernel menunjukkan akurasi yang sama besar yaitu 100%.

- Hasil Pengujian Akurasi Kedua  
Berikut hasil pengujian kedua dengan menggunakan banyaknya data latih 200 dan

data uji 100, nilai akurasi yang didapatkan bisa dilihat pada Tabel 6 berikut:

**Tabel 6 Hasil Pengujian Akurasi Kedua**

Kondisi	Linear	RBF		
		$\gamma=1$	$\gamma=2$	$\gamma=3$
SVM	96%	94%	91%	89%
		Polynomial		
		$n=1$	$n=2$	$n=3$
		73%	74%	74%

Hasil pengujian tersebut menunjukkan bahwa metode *Support Vector Machine* kernel linear menunjukkan akurasi yang paling besar dengan nilai akurasi 96% dibandingkan dengan RBF dan Polynomial. Dari hasil pengujian akurasi pertama 96 data uji terprediksi secara benar dan 4 data uji terprediksi secara salah.

### 3. PENUTUP

#### 3.1 Kesimpulan

Berdasarkan hasil implementasi dan pengujian *Support Vector Machine* pada konten instagram berdasarkan komentar ini telah dapat memenuhi tujuan dalam penelitian yaitu mengetahui besar

akurasi yang didapatkan dengan mengimplementasikan metode *Support Vector Machine* untuk klasifikasi konten instagram berdasarkan komentar dengan besar akurasi yang didapatkan untuk pengujian pertama sebesar 100% dan pengujian kedua sebesar 96%. Maka dapat ditarik kesimpulan klasifikasi konten instagram berdasarkan komentar dengan metode *Support Vector Machine* dengan Kernel *Linear* dari pengujian pertama dan kedua dapat dilakukan dengan menggunakan metode *Support Vector Machine*. Dari hasil tersebut dapat dilihat bahwa metode *Support Vector Machine* untuk mengklasifikasikan konten Instagram berdasarkan komentar dapat digunakan.

#### 3.2 Saran

Berdasarkan hasil penelitian dalam tahapan *preprocessing* data latih dan data uji untuk komentar dalam postingan instagram perlu diperhatikan, terutama pada komentar yang tidak baku, bahasa asing dan bahasa daerah. Karena masih ada kata yang tidak sesuai dengan Kamus Besar Bahasa Indonesia(KBBI), maka kata-kata tersebut perlu diubah agar mudah di proses ketika diklasifikasikan.

### DAFTAR PUSTAKA

- [1] B. A. Kuncoro, "TF-IDF Method in Ranking Keywords of Instagram Users ' Image Captions," pp. 1–5, 2015.
- [2] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. SAINS DAN SENI ITS Vol. 4, No.2*, vol. 4, no. 2, pp. 248–253, 2015.
- [3] K. F. Siti Nur Asiyah, "Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K- Nearest Neighbor," *J. SAINS dan SENI ITS*, vol. 5, no. 2, pp. 317–322, 2016.
- [4] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, "Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," *J. Infotel*, vol. 9, no. 4, p. 416, 2017.
- [5] I. Adiwijaya, "Text Mining dan Knowledge Discovery," *Kolok. bersama komunitas datamining Indones. soft-computing Indones.*, pp. 1–9, 2006.
- [6] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, pp. 4704–4713, 2018.
- [7] A. A. Amrullah, A. Tanton, N. Hamdani, R. T. R. L. Bau, M. R. Ahsan, and E. Utami, "Reviewatas Analisis Sentimen Pada Twitter Sebagai Representasi Opini Publik Terhadap Bakal Calon Pemimpin," *Pros. Semin. Nas. Multi Disiplin Ilmu Call Pap. Unisbank*, vol. 2, no. 1, pp. 978–979, 2016.
- [8] L. Sofiyana, Z. Abidin, and H. Nurhayati, "Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Dengan Menggunakan K-Nearest Neighbor," vol. 1, no. January, pp. 194–299, 2012.
- [9] I. H. Witten, *11 - AAI*. 2002.
- [10] A. Purwarianti, A. Andhika, A. F. Wicaksono, I. Afif, and F. Ferdian, "InaNLP: Indonesia natural language processing toolkit, case study: Complaint tweet classification," *4th IGNITE Conf. 2016 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2016*, pp. 5–9, 2016.
- [11] F. Z. Tala, "A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia," *J. Teknosains*, vol. 6, no. 2, p. 113, 2017.
- [12] D. S. Harjanto, S. N. Endah, and N. Bahtiar, "Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frecency Invers Document Frequency (TF-IDF)," *J. Sains dan Mat.*, vol. 20, no. 3, pp. 64–70, 2012.
- [13] E. Rainarli and A. Romadhan, "Perbandingan Simple Logistic Classifier dengan Support Vector Machine dalam Memprediksi Kemenangan Atlet," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 3, no. 2, p. 87, 2017.
- [14] N. Christianini, "Support Vector and Kernel Machines." 2001.
- [15] M. H, "Support Vector Machines-Kernels

and The Kernel Trick,” *An Elabor. Hauptseminar Read. Club Support Vector Mach.*, 2006.

[16] Eko Prasetyo, “DATA MINING - Mengolah

Data menjadi Informasi Menggunakan Matlab,” *Yogyakarta: ANDI*, 2014.