

BAB 2

LANDASAN TEORI

2.1 Roblox

Roblox adalah platform dan etalase *online* tempat pengguna bermain *game*. *Roblox* lebih mirip dengan platform Steam karena terdapat banyak *game* yang dibuat oleh pengembang lain atau singkatnya oleh pengguna itu sendiri. *Roblox* memiliki keunggulan dibanding *game* lainnya, yaitu pengguna bisa membuat *game* di dalamnya. Dalam hal ini *game-game* tidak dikembangkan secara resmi oleh *Roblox Corporation*, tetapi oleh pengguna platform. Saat ini, situs web resmi membanggakan bahwa penggunanya telah menerbitkan lebih dari 20 juta *game* di platform [8]. CEO *Roblox*, Dave Baszucki ingin menghuni *metaverse*, artinya tempat *online* yang dimana tempat untuk bekerja, bermain dan menghibur diri. Menurutnya *Roblox* kemungkinan akan menjadi salah satu kandidat paling kredibel untuk membangun *metaverse*, alam semesta dunia virtual yang semuanya berhubungan, seperti dalam novel *Snow Crash*. Kata Baszucki, dalam *metaverse* sendiri, *Roblox* memiliki delapan fitur *metaverse*, yaitu identitas, variasi, teman, di mana saja, imersif, ekonomi, gesekan rendah, dan kesopanan [9].

2.2 Google Play

Google Play Store merupakan layanan digital yang didalamnya mencakup toko daring seperti buku, film, *games* dan aplikasi lainnya. *Google Play Store* juga bisa sebagai sarana timbal balik, misalkan aplikasi atau *software* yang kita buat dipublikasikan di *Play Store*, kita bisa melihat banyak ulasan dari pengguna yang telah menggunakan aplikasi, seberapa efisien kah, apakah ada *bug*. Kita bisa melihat timbal baliknya dari hal tersebut [10].

Sebagian besar aplikasi di *Google Play Store* mendapat peringkat lebih tinggi, yang menimbulkan kepercayaan di antara pengguna. Sesuai laporan yang diterbitkan *Statista*, pada Januari 2022, ada sekitar 142.368 aplikasi di *Google Play Store* dengan peringkat rata-rata 4,5 dari 5 bintang. Secara total 1,44 juta aplikasi memiliki peringkat dan 2,13 juta aplikasi lainnya di *Google Play Store* memiliki kurang dari tiga peringkat pengguna. Dalam hal kategori, aplikasi *game* adalah kategori aplikasi paling populer di *Google Play Store*, terhitung 13,72 persen dari aplikasi yang tersedia di seluruh dunia [11]. Pada penelitian ini, akan diambil data ulasan para pengguna mengenai aplikasi *game Roblox*.

2.3 Analisis Sentimen

Analisis Sentimen adalah salah satu analisis utama yang saat ini digunakan dengan tujuan untuk mengklasifikasikan sentimen dan opini yang dihasilkan oleh makhluk dan dalam teks. Teknik *Machine Learning* Sebagian besar digunakan dalam analisis sentimen [12]. Singkatnya, analisis sentimen atau *opinion mining* bertujuan untuk mengidentifikasi ulasan positif dan negatif yang tersirat dalam teks [3]. Seiring pertumbuhan sosial media, blog dan forum memungkinkan orang terus-menerus berbagi komentar dan ide pada topik yang berbeda. Misalnya, dalam membeli produk konsumen, tidak perlu bertanya kepada teman dan keluarga, karena banyak ulasan dan diskusi tersedia *online* pada produk tertentu. Demikian pula untuk sebuah perusahaan atau organisasi, kita tidak perlu lagi melakukan survei untuk memahami opini publik tentang layanan mereka sebagai sosial media untuk membantu mengatur ulang bisnis mereka [13].

2.3.1. Analisis Sentimen Berbasis Aspek

Analisis sentimen berbasis aspek merupakan salah satu cara menganalisis kata-kata pada bagian-bagian tertentu [14]. Ada 2 proses analisis sentimen berbasis aspek, yaitu:

1. *Aspect Extraction*

Aspect Extraction ialah proses mengekstraksi semua istilah yang terkait dengan aspek produk/layanan, seperti Bahasa gaul, singkatan, *emoticon*, dan kesalahan tipografi.

2. *Aspect Classification*

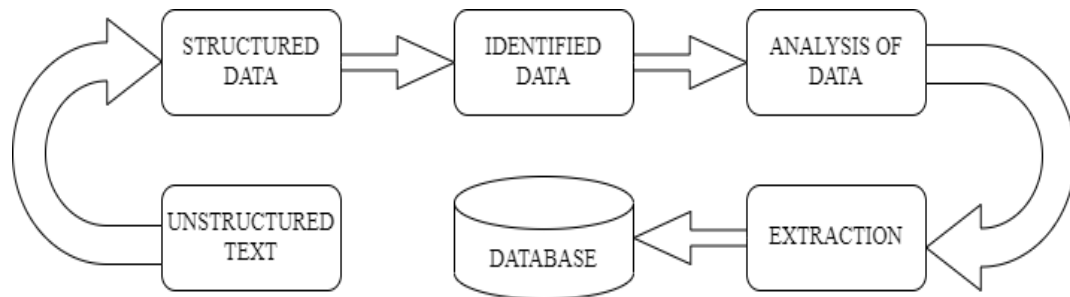
Aspect Classification adalah proses mengklasifikasikan sentimen dengan 2 kategori, yaitu positif dan negatif.

2.4 Text Mining

Text mining merupakan proses untuk mengekstraksi pola yang menarik dan signifikan untuk mengeksplorasi pengetahuan dari sumber data tekstual. *Text mining* ialah bidang multidisiplin yang didasarkan pada *information retrieval*, *data mining*, *machine learning*, *statistics*, dan *computational linguistic*. Ada 5 tahapan dalam *text mining* di bawah ini:

Langkah-langkah *Text Mining*:

1. Mengumpulkan informasi dari data yang tidak terstruktur
2. Ubah informasi yang diterima ini menjadi data terstruktur
3. Mengidentifikasi pola dari data terstruktur
4. Analisis polanya
5. Ekstrak informasi berharga dan simpan di basis data



Gambar 2.1 *Text Mining* [15]

Kemudian teknologi yang digunakan oleh *text mining* seperti *information extraction*, *summarization*, *categorization*, *clustering*, dan *information visualization* [15].

2.5 Preprocessing

Text Preprocessing adalah proses menyiapkan teks dokumen atau kumpulan data mentah. Tujuan *text processing* adalah mengubah data teks yang awalnya tidak terstruktur menjadi data yang terstruktur [16]. Tahapan dalam *preprocessing* adalah sebagai berikut:

2.5.1 *Cleaning*

Cleaning, yaitu proses menghilangkan kata-kata selain huruf seperti nomor, simbol, tanda baca, emoji, dll [17].

2.5.2 *Case Folding*

Case folding, yaitu proses mengubah kata *uppercase* (huruf kapital) menjadi *lowercase* (huruf kecil). [16].

2.5.3 *Tokenizing*

Tokenizing, yaitu memisahkan kalimat menjadi kata tunggal [17].

2.5.4 *Stopword Removal*

Stopword Removal, merupakan proses untuk menghilangkan kata-kata yang dianggap tidak berpengaruh terhadap kalimat seperti kata “di”, “dari”, “yang”, dll [17].

2.5.5 Stemming

Stemming, merupakan proses untuk mengubah kata-kata yang ada menjadi bentuk kata dasar seperti kata “kesalahan” menjadi “salah” [17].

2.6 Pembobotan TF-IDF

Metode TF-IDF adalah suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen [18]. Berikut rumus untuk TF-IDF.

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (2.1)$$

$$idf_t = \log\left(\frac{D}{df_t}\right) \quad (2.2)$$

$$W_{d,t} = tf_{d,t} \times IDF_{d,t} \quad (2.3)$$

Keterangan:

tf = banyaknya kata yang dicari pada sebuah dokumen

$\max(tf)$ = jumlah kemunculan banyaknya term pada dokumen yang sama.

Nilai D = total dokumen

df_t = jumlah dokumen yang mengandung term t .

IDF = Inversed Dokument Frequency ($\log_2(D/df)$)

d = dokumen ke- d

t = kata ke- t dari kata kunci

W = bobot dokumen ke- d terhadap kata ke- t

2.7 Support Vector Machine

Support Vector Machine adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang berdasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik. 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 dan lain sebagainya [19]. Klasifikasi data menggunakan *Support Vector Machine* dapat dibedakan atas 2 berdasarkan kernel yang digunakan [20], yaitu:

1. SVM Linear

Data yang akan digunakan dengan metode SVM akan diberi variabel dengan $x_i \in R^d$ dan untuk label aktual akan diberi variabel dengan $y_i \in \{1,2\}$ dengan $i = 1,2$, dst. Pada tiap data akan diberi label dengan notasi 1 sebagai *sentiment* positif, 0 sebagai sentimen netral dan -1 sebagai label *actual sentiment negative*. Dengan begitu, tiap data dan label akan diberi variabel dengan $y_i \in \{-1,0,1\}$. Dengan demikian, *Support Vector Machine* dapat diberikan persamaan terhadap $y_i = -1, y_i = 0, y_i = 1$ dalam persamaan 2.4, 2.5, 2.6

$$X_i * w + b \geq 1 \quad (2.4)$$

$$X_i * w + b = 0 \quad (2.5)$$

$$X_i * w + b \geq -1 \quad (2.6)$$

Keterangan:

W = normal bidang

b = penentuan bidang dalam koordinat

Dengan mengetahui garis normal dan posisi terhadap bidang yang *relative* terhadap titik koordinat maka kita akan menentukan nilai jarak terdekat terhadap margin. Hal tersebut menjadikan sebagai pembatas atau *support* untuk menentukan perbedaan terhadap

nilai sentimen -1 atau negatif, 0 atau netral dan 1 atau positif. Nilai jarak margin dapat dioptimalkan dengan menggunakan *hyperplane* dan titik terdekatnya dengan $1/|w|$. Untuk mendapatkan titik tersebut perlu adanya persamaan *Quadratic Programming* (QP) *Problem*, dengan mencari persamaan terhadap nilai titik minimum pada persamaan 2.7 dan mendapatkan nilai *constraint* dengan persamaan 2.8.

$$\min_{\frac{1}{2}} ||w||^2 \quad (2.7)$$

$$y_i(wx_i + b) \geq 1, i = 1, \dots, n \quad (2.8)$$

Untuk mengoptimisasi pada linear dan tidak membatasi masalah permodelan persamaan dapat diberikan *lagrange Multiplier*. Dengan adanya fungsi tujuan, fungsi *langragian* fungsi kendala dapat mempresentasikan kondisi yang membatasi untuk meningkatkan optimasi. Dengan menambahkan metode *lagrange Multiplier* persamaan dapat menjadi 2.9

$$\min_{w,b} L_p(w, b, a) = \frac{1}{2} |w|^2 - \sum_{i=1}^n a_i y_i (x_i \cdot w + b) + \sum_{i=1}^n a_i \quad (2.9)$$

w dan b memberi nilai faktor terhadap L_p maka, perlu diperhatikan untuk meminimalkan nilai L_p dengan tambahan kontrain $a_i \geq 0$ maka ketika $\frac{\delta}{\delta b} L_p(w, b, a) = 0$ diperoleh persamaan 2.10, dan terhadap nilai w ketika $\frac{\delta}{\delta w} L_p(w, b, a) = 0$ maka diperoleh persamaan 2.11 berikut.

$$\sum_{i=1}^n a_i y_i = 0 \quad (2.10)$$

$$\sum_{i=1}^n a_i y_i x_i \quad (2.11)$$

Namun terdapat kekurangan yang umumnya w dapat bernilai ∞ atau tak terhingga sehingga tidak mendapatkan hasil yang tepat berdasarkan support vector yang diberikan perlu adanya persamaan *problem L_D* sebagai substitusi terhadap *problem L_p* . Subtitusi persamaan akan memperoleh dua *problem L_D* seperti pada persamaan 2.12

$$L_D(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1, j=1}^n a_i a_j y_i y_j x_i x_j \quad (2.12)$$

Dengan persamaan 2.11 kita dapat menentukan *constraint* $\min_{w,b} Lp = \max_a L_d$ sehingga untuk menentukan margin terhadap hasil nilai $\max_a L_d$ dapat ditentukan sebagai persamaan 2.13

$$\max_a L_d \equiv \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1, j=1}^n a_i a_j y_i y_j x_i x_j \quad (2.13)$$

Maka persamaan tersebut dapat menentukan *support vector* yang jelas untuk membagi berdasarkan nilai masing-masing terhadap w dan b sehingga pemisah nilai akan menjadi faktor untuk menentukan hasil sentiment berdasarkan perhitungan persamaan diatas. Nilai support vector nantinya akan menjadi pemisah berdasarkan $a \geq 0$ yang menjadi hyperplane dan $\sum_{i=1}^n a_i y_i = 0$ menjadi menghasilkan nilai a_i dengan nilai 0

2. SVM Nonlinear

Support Vector Machine Non-Linear merupakan cara untuk memisahkan data-data dari yang sifatnya non-linear menjadi linear dengan menggunakan fungsi *kernel*. Fungsi *kernel* dalam metode ini juga disebut *kernel trick*. *Kernel trick* merupakan fungsi yang mengelompokkan data dari dimensi rendah menjadi data dimensi tinggi. Terdapat beberapa macam fungsi *kernel Non-Linear* yang dapat digunakan pada sebuah aplikasi untuk mengatasi masalah, seperti:

a) *Kernel Polynomial*

$$K(x, y) = (x \cdot y + 1)^p \quad (2.14)$$

b) *Kernel Gaussian RBF*

$$K(x, y) = e^{-\|x-y\|^2/2\sigma^2} \quad (2.15)$$

c) *Kernel Sigmoid*

$$K(x, y) = \tanh(kx \cdot y - \delta) \quad (2.16)$$

Keterangan:

p : pangkat (*degree of*)

σ : nilai *sigma*

δ : nilai *delta*

Berikut merupakan Langkah-langkah proses klasifikasi menggunakan algoritma *Support Vector Machine* [20], yaitu:

1. Menghitung *Kernel SVM*

Proses pertama yaitu menentukan dot *product* setiap data dengan memasukkan fungsi *kernel* baik *kernel* Linear maupun *kernel* Non-Linear

2. Menghitung nilai matriks *Hessian*

Fungsi dari matriks *Hessian* adalah untuk mengidentifikasi optimum relatif suatu nilai fungsi. Untuk menghitung matriks *Hessian*, dapat ditunjukkan pada persamaan (2.17).

$$[D]_{ij} = y_i y_j (K(x, y) + \lambda^2) \quad (2.17)$$

Keterangan:

$[D]_{ij}$: nilai matriks *Hessian* ke- i, j

y_i : kelas data ke- i

y_j : kelas data ke- j

λ : nilai *lambda* yang diinisialisasi

3. Menghitung nilai *gamma*

Nilai *gamma* (γ) berfungsi untuk mempercepat fungsi pada kernel SVM untuk mendapatkan akurasi classifier model klasifikasi yang optimal. Perhitungan nilai *gamma* dapat ditunjukkan pada persamaan (2.18).

$$\gamma = \frac{cLR}{\max(Hessian)} \quad (2.18)$$

Keterangan:

cLR : nilai *constant Learning Rate*

4. Proses perhitungan *Sequential Training SVM*, terdapat beberapa tahapan yaitu:

a. Hitung nilai *error* (E_i)

$$E_i = \sum_j^n = 1 \alpha_i D_{ij} \quad (2.19)$$

b. Hitung nilai $\delta\alpha_i$

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \quad (2.20)$$

c. Hitung nilai α_i

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (2.21)$$

Keterangan:

α_i : nilai parameter *alpha* data ke-*i*

C : nilai parameter *complexity*

$\delta\alpha_i$: nilai delta alpha data ke-*i*

5. Menghitung nilai bobot/*weight* (w)

Sebelum menghitung nilai w , terlebih dahulu harus menentukan nilai α_i *max* kelas positif dan nilai α_i *max* kelas negatif. Pada proses perhitungan nilai w dapat ditunjukkan pada persamaan berikut.

$$w.x^+ = K(x.y^+) * \max(\alpha_i^+) * yi \quad (2.22)$$

$$w.x^- = K(x.y^+) * \max(\alpha_i^+) + yi \quad (2.23)$$

Keterangan:

$w.x^+$: nilai bobot kelas positif

$w.x^-$: nilai bobot kelas *negative*

$K(x.y^+)$: kernel kelas α_i *max* positif

$K(x.y^-)$: kernel kelas α_i *max* negative

6. Menghitung nilai bias (b)

Proses perhitungan nilai *bias* dapat ditunjukkan pada Persamaan (2.24).

$$b = -\frac{1}{2} (\sum w.x^+ + \sum w.x^-) \quad (2.24)$$

7. Proses Perhitungan Pengujian SVM

Pada proses perhitungan pengujian SVM terdapat beberapa tahapan, yaitu:

a. Perhitungan *kernel* data *testing*

Setelah mendapat nilai α , w , dan b dari proses training SVM maka Langkah pertama dalam proses pengujian adalah menghitung semua *dot product kernel* antara data *testing* dan data *training* berdasarkan perhitungan *kernel* yang digunakan sebelumnya pada proses *training* SVM. Selanjutnya dilakukan perhitungan $\sum \alpha y_i (K(x, y))$ terhadap keseluruhan data *testing* menggunakan perhitungan pada Persamaan (2.25).

$$\sum \alpha_i * y_i * (K(x, y)) \quad (2.25)$$

b. Melakukan perhitungan $f(x)$

Perhitungan nilai $f(x)$ bertujuan untuk mencari nilai klasifikasi dari setiap data *testing* dengan menggunakan fungsi *sign*. Perhitungan fungsi klasifikasi dapat ditunjukkan pada Persamaan (14).

$$f(x) = \text{sign} (\sum_{i=1}^m \alpha_i y_i K(x, y) + b) \quad (2.26)$$

2.8 Confusion Matrix

Untuk mengetahui performansi dari penelitian ini, akan dilakukan empat evaluasi terpisah untuk masing-masing tahap [23]. Evaluasi tersebut yaitu *accuracy*, *precision*, *recall*, dan *f1-score*. Rumus-rumus keempat evaluasi tersebut yaitu:

$$\text{Akurasi} = \frac{\text{Jumlah aspek atau kategori yang benar dideteksi}}{\text{jumlah aspek atau kategori yang beranotasi}} \quad (2.27)$$

TP (*True Positive*) melambangkan nilai positif yang sesuai prediksi, FN (*False Negative*) melambangkan nilai positif yang dianggap negatif, FP (*False Positive*) melambangkan nilai negatif yang dianggap positif, dan TN (*True Negative*) melambangkan nilai negatif yang sesuai prediksi.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.28)$$

Tabel 2.1 *Confusion Matrix*

		Prediksi (Hasil Ekstraksi Aspek)	
		Diekstrak	Tidak Diekstrak
Aktual (<i>Dataset</i> berlabel)	Relevan	(<i>True Positive</i>)	TN (<i>True Negative</i>)
		Jumlah aspek yang terdeteksi sesuai data aktual	Jumlah aspek yang terlabel, tapi tidak diekstrak
	Tidak Relevan	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)
		Jumlah aspek yang tidak terlabel, tapi diekstrak	Jumlah aspek yang tidak terlabel dan tidak diekstrak

Akan didapatkan nilai *precision*, yaitu jumlah positif yang sesuai dibagi dengan jumlah positif yang didapat oleh hasil klasifikasi

$$recall = \frac{TP}{TP+FN} \quad (2.29)$$

Akan didapatkan juga nilai *recall* yaitu jumlah sesuai dibagi dengan jumlah data yang ditemukan oleh hasil klasifikasi.

$$F1 = 2x \frac{precision \times recall}{precision + recall} \quad (2.30)$$

Akan dihitung nilai *f1-score* sesuai perhitungan untuk mencari keseimbangan antara *precision* dan *recall*.

2.9 Penelitian-Penelitian Terkait

Tabel 2.2 Penelitian-Penelitian Terkait

Review Literatur Pertama [6]	
Judul Artikel	Perbandingan Metode Klasifikasi <i>Support Vector Machine</i> dan <i>Naïve Bayes</i> untuk Analisis Sentimen pada Ulasan Tekstual di <i>Google Play Store</i>
Penulis	Luthfi Budi Ilmawan, Muhammad Aliyazid Mude
Judul Jurnal/ <i>Proceeding</i>	ILKOM Jurnal Ilmiah
Tahun Penerbitan	2020
Masalah Utama yang diangkat	Mengukur kinerja klasifikasi <i>Support Vector Machine</i> dibandingkan dengan klasifikasi <i>Naïve Bayes</i> untuk proses Analisis Sentimen tekstual berbahasa Indonesia pada <i>Google Play Store</i>
Metode Ekstraksi	<i>Case folding, stopword and punctuation elimination, negation tag</i>
Metode Klasifikasi	<i>Support Vector Machine</i> , dan <i>Naïve Bayes</i>
Hasil Penelitian dan Kesimpulan	<p>Hasil: dari 606 data testing, rata-rata akurasi <i>Naïve Bayes</i> classifier yaitu 75,41%, sedangkan akurasi dari SVM classifier mencapai 81,46%</p> <p>Kesimpulan: Hasil akurasi dari <i>Support Vector Machine classifier</i> memiliki nilai yang lebih tinggi jika dibandingkan dengan akurasi dari <i>Naïve Bayes classifier</i> untuk mengklasifikasi ulasan tekstual berbahasa Indonesia pada <i>Google Play Store</i></p>
Review Literatur Kedua [7]	
Judul Artikel	Perbandingan <i>Naïve Bayes</i> , SVM, dan k-NN untuk Analisis Sentimen <i>Gadget</i> Berbasis Aspek
Penulis	Jessica Widyadhana Iskandar, Yessica Nataliani
Judul Jurnal/ <i>Proceeding</i>	JURNAL Rekayasa Sistem dan Teknologi Informasi
Tahun Penerbitan	2021

Masalah Utama yang diangkat	Membandingkan <i>Naïve Bayes</i> , SVM, dan k-NN dengan tingkat <i>accuracy</i> , <i>recall</i> , <i>precision</i> , dan <i>F-measure</i> terbaik dalam klasifikasi komentar Samsung Galaxy Z Flip 3
Metode Ekstraksi	<i>Transform Case</i> , <i>Tokenizing</i> , <i>Normalization</i> , <i>Stemming</i>
Metode Klasifikasi	Naïve Bayes, SVM, dan k-NN
Hasil Penelitian dan Kesimpulan	<p>Hasil: <i>accuracy</i> dari masing-masing model klasifikasi yaitu NB sebesar 83.54%, SVM sebesar 96.43%, dan k-NN sebesar 59.68%. Kemudian rata-rata <i>precision</i>, yaitu NB sebesar 81.93%, SVM sebesar 96.22%, dan k-NN sebesar 56.43%. Hasil <i>F-Measure</i>, NB sebesar 84.03%, SVM sebesar 97.25%, dan k-NN sebesar 74.69%. Hasil <i>recall</i>, k-NN memiliki rata-rata lebih unggul.</p> <p>Kesimpulan: model klasifikasi SVM menunjukkan hasil terbaik dari hasil analisis dan pengujian yang telah dilakukan terhadap komentar Youtube <i>Gadget</i> Samsung Galaxy Z Flip 3 dalam aspek harga, spesifikasi, dan citra merek.</p>
Review Literatur Ketiga [16]	
Judul Artikel	Analisis Sentimen Pengguna Aplikasi <i>Marketplace</i> Tokopedia Pada Situs <i>Google Play</i> Menggunakan Metode <i>Support Vector Machine</i> (SVM), <i>Naïve Bayes</i> , dan <i>Logistic Regression</i>
Penulis	Meishita Inelza Putri, Iqbal Kharisudin
Judul Jurnal/ <i>Proceeding</i>	PROSIDING SEMINAR NASIONAL MATEMATIKA
Tahun Penerbitan	2022
Masalah Utama yang diangkat	Kebanyakan data <i>review</i> cenderung tidak seimbang (<i>imbalanced datasets</i>) dari segi jumlah tiap kelas individu, misalnya cenderung ke arah kelas positif atau sebaliknya pada aplikasi Tokopedia.
Metode Ekstraksi	<i>Web scraping</i>

Metode Klasifikasi	<i>Support Vector Machine, Naïve Bayes, Logistic Regression</i>
Hasil Penelitian dan Kesimpulan	<p>Hasil: Secara keseluruhan, metode SVM merupakan metode yang tepat untuk mengklasifikasikan data <i>review</i> pengguna aplikasi Tokopedia dibandingkan dengan <i>Naïve Bayes, Logistic Regression</i>. Alasannya karena metode tersebut memiliki nilai AUC sebesar 0,9545 yang menunjukkan bahwa metode ini memiliki kinerja klasifikasi yang sangat baik dan penggunaan SMOTE dapat meningkatkan akurasi model pada data <i>imbalanced</i> hasil <i>review</i> pengguna <i>marketplace</i> Tokopedia</p> <p>Kesimpulan: Metode klasifikasi yang tepat digunakan untuk klasifikasi <i>review</i> pengguna Tokopedia adalah <i>k-fold cross validation</i> dengan SMOTE, yaitu metode klasifikasi <i>Support Vector Machine</i> dengan nilai AUC sebesar 1,000, sehingga metode tersebut tergolong memiliki kinerja klasifikasi sangat baik.</p>
<i>Review Literatur Keempat [17]</i>	
Judul Artikel	Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma <i>Support Vector Machine</i>
Penulis	Muhammad Diki Hendriyanto, Azhari Ali Ridha, Ultach Enri
Judul Jurnal/ <i>Proceeding</i>	INTECOMS
Tahun Penerbitan	2022
Masalah Utama yang diangkat	Ulasan di <i>Google Play Store</i> cukup banyak dan tidak terstruktur
Metode Ekstraksi	<i>Cleaning, Case Folding, Tokenizing, Normalization, Stopword Removal, Stemming</i>
Metode Klasifikasi	<i>Support Vector Machine</i>
Hasil Penelitian dan Kesimpulan	Hasil: dengan menggunakan <i>kernel RBF</i> , hasil akurasi sebesar 92,31%, <i>precision</i> 96,3%, <i>recall</i> 89,66%, <i>f1-score</i> 92,86%

	<p>Kesimpulan: Pada hasil evaluasi kinerja empat kernel algoritma <i>Support Vector Machine</i> dalam menganalisis sentimen ulasan aplikasi MOLA dengan tiga skenario <i>split data</i> diperoleh hasil terbaik skenario 1 dengan rasio perbandingan <i>data training</i> dan <i>data testing</i> 10% dengan hasil <i>accuracy</i> 92,31%, <i>precision</i> 96,3%, <i>recall</i> 89,66%, dan <i>f1-score</i> 92,86%.</p>
<i>Review Literatur Kelima [19]</i>	
Judul Artikel	Penerapan Metode <i>Support Vector Machine</i> Pada Diagnosa Hepatitis
Penulis	Raudlatul Munawarah, Oni Soesanto, M. Reza Faisal
Judul Jurnal/ <i>Proceeding</i>	Kumpulan Jurnal Ilmu Komputer
Masalah Utama yang diangkat	Mendiagnosis seseorang mengidap penyakit tertentu atau tidak berdasarkan hasil tes fungsi hati atau rekam medis pasien, khususnya penyakit hepatitis
Metode Ekstraksi	-
Metode Klasifikasi	<i>Support Vector Machine</i>
Hasil Penelitian dan Kesimpulan	<p>Hasil: data yang berjumlah 579 terbagi menjadi dua untuk <i>data training</i> dan <i>data testing</i>. <i>Data training</i> menggunakan 100 data kelas label positif dan 100 data negatif. Sisa data digunakan untuk <i>testing</i>. Hasil akhir menunjukkan bahwa pasien tersebut positif mengidap hepatitis.</p> <p>Kesimpulan: Metode <i>Support Vector Machine</i> dapat digunakan untuk mendiagnosis penyakit hepatitis dengan belajar dari pembelajaran yang dihasilkan dari pelatihan <i>data training</i>.</p>
<i>Review Literatur Keenam [1]</i>	

Judul Artikel	Menganalisis Informasi <i>Metaverse</i> Pada Game Online <i>Roblox</i> Secara Garis Besar
Penulis	Ii Sopiandi, Deffy Susanti
Judul Jurnal/ <i>Proceeding</i>	Jurnal PETISI
Tahun Penerbitan	2022
Masalah Utama yang diangkat	-
Metode Ekstraksi	-
Metode Klasifikasi	-
Hasil Penelitian dan Kesimpulan	<p>Hasil: Informasi <i>Game Roblox</i>, Pengguna bisa berkomunikasi di dunia virtual <i>metaverse</i>, dan Peluang bisnis <i>voucher games</i>.</p> <p>Kesimpulan: Di dunia <i>games</i> sekarang tidak ada yang tidak mungkin kita bisa melakukan banyak hal, selain kita bisa bermain <i>games</i> di dalamnya kita bisa ikut andil dalam penciptaan <i>games</i> sendiri. Dengan adanya <i>metaverse</i> orang orang bisa membangun jaringan di dunia <i>games</i> dan berkomunikasi secara virtual hologram dan membangun bisnis sendiri dalam menghasilkan uang</p>
<i>Review literatur Ketujuh [21]</i>	
Judul Artikel	Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode <i>Support Vector Machine</i> (SVM)
Penulis	Salsabila Rahma Yustihan, Putra Pandu Adikara, Indriati
Judul Jurnal/ <i>Proceeding</i>	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer
Tahun Penerbitan	2021
Masalah Utama yang diangkat	Banyaknya ulasan restoran di internet, pembaca akan kesulitan menemukan informasi yang relevan terhadap preferensi mereka selain kategori aspek dan sedikitnya situs web yang menyediakan sistem penilaian berdasarkan aspek.

Metode Ekstraksi	<i>TF-IDF</i>
Metode Klasifikasi	<i>Support Vector Machine</i>
Hasil Penelitian dan Kesimpulan	<p>Hasil: Hasil pengujian klasifikasi aspek menunjukkan bahwa sistem sudah cukup mampu dalam mengidentifikasi kelas yang bukan termasuk aspeknya dengan benar namun sistem belum cukup berhasil untuk mengidentifikasi semua kelas yang termasuk aspeknya dengan benar. Hal ini juga berlaku pada aspek lain dikarenakan jumlah data saat <i>training</i> tidak seimbang yang mana jumlah data non-aspek jauh lebih besar daripada jumlah data pada aspek tersebut.</p> <p>Kesimpulan: hasil evaluasi yang diperoleh oleh metode SVM dalam mengklasifikasikan aspek dengan menggunakan pendekatan <i>macro averaging</i> menghasilkan <i>precision</i> sebesar 0,94, <i>recall</i> sebesar 0,6, <i>accuracy</i> sebesar 0,88, dan <i>f-measure</i> sebesar 0,73. Hasil evaluasi tersebut menunjukkan bahwa sistem sudah cukup mampu dalam mengidentifikasi kelas yang mempunyai sentimen positif dengan benar, namun belum berhasil mengidentifikasi kelas yang mempunyai sentimen negatif dikarenakan jumlah data <i>training</i> yang digunakan saat proses pelatihan klasifikasi sentimen tidak seimbang di mana data positif lebih banyak dibandingkan dengan data negatif</p>
<i>Review Literatur Kedelapan [4]</i>	
Judul Artikel	Facet Analysis of Video Game Genres
Penulis	Jin Ha Lee, Natascha Karlova, Rachel Ivy Clarke, Katherine Thornton, Andrew Perti
Judul Jurnal/ <i>Proceeding</i>	IConference 2014 Proceedings
Tahun Penerbitan	2014

Masalah Utama yang diangkat	Desainer <i>game</i> , produsen, cendekiawan, pendidik, pemain, dan orang tua muda gamer semua membutuhkan cara yang berarti untuk menemukan, mengakses, dan menafsirkan video game. Sebagai langkah awal menyediakan akses yang kuat ke <i>video game</i> untuk beragam pemangku kepentingan, kita perlu memahami informasinya diberikan kepada pengguna melalui akses <i>video game</i> saat ini.
Metode Ekstraksi	-
Metode Klasifikasi	-
Hasil Penelitian dan Kesimpulan	<p>Hasil: Dua belas aspek diidentifikasi, masing-masing mewakili karakteristik divisi yang berbeda terkait dengan <i>video game genre</i>.</p> <p>Kesimpulan: Makalah ini melaporkan langkah pertama dalam memahami kompleksitas genre <i>video game</i> untuk merancang sebuah skema yang lebih kuat untuk mewakili informasi ini. Analisis label <i>game</i> yang digunakan dalam <i>game</i> terkait situs web dan katalog mengungkapkan bahwa elemen <i>metadata</i> "genre" sangat kelebihan beban dengan banyak dimensi informasi. Melalui metode analisis segi, dimungkinkan untuk memastikan dan mewakili berbagai jenis informasi yang tertanam dalam label genre <i>video game</i> saat ini secara fleksibel dan dapat diperluas cara.</p>
<i>Review Literatur Kesembilan [23]</i>	
Judul Artikel	Analisis Sentimen Berbasis Aspek Terhadap Ulasan Restoran Berbahasa Indonesia menggunakan <i>Support Vector Machine</i>
Penulis	Tri Jaka Pamungkas, Ade Romadhony
Judul Jurnal/ <i>Proceeding</i>	Jurnal Gaussian
Tahun Penerbitan	2021

Masalah Utama yang diangkat	Dalam suatu ulasan dapat berisi beberapa aspek yang masing-masing mengandung polaritas yang berbeda, sehingga akan membuat sebuah kalimat tidak sepenuhnya hanya mengandung satu jenis polaritas.
Metode Ekstraksi	<i>Grammar rules</i>
Metode Klasifikasi	Support Vector Machine
Hasil Penelitian dan Kesimpulan	<p>Hasil: evaluasi ekstraksi aspek dengan menggunakan <i>dependency rule</i> dan leksikal opini Bahasa Indonesia dengan hasil paling baik dengan nilai <i>precision</i> 39,195% dan nilai <i>recall</i> 40,636%. Kemudian pada tahap klasifikasi polaritas aspek dengan pembobotan berdasarkan leksikal opini Bahasa Indonesia dapat diketahui dengan akurasi paling baik sebesar 38.352%. kalimat dengan metode SVM, dapat diketahui hasil rata-rata F1-Score mendapatkan nilai sebesar 68%. Terakhir pada tahap klasifikasi polaritas kategori, dapat diketahui nilai akurasi paling baik sebesar 15.119%.</p> <p>Kesimpulan: Hasil yang paling baik didapatkan pada pengujian dengan melakukan <i>stemming</i> tanpa menghapus tanda baca. Secara keseluruhan, sistem tidak dapat bekerja secara optimal akibat dari ketidaksesuaian bentuk data terhadap sistem yang dibangun. Untuk kedepannya, lebih baik untuk diadakannya filterisasi terhadap <i>dataset</i> yang akan sesuai dengan sistem yang dibangun.</p>
<i>Review Literatur Kesepuluh [24]</i>	
Judul Artikel	Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Terhadap Kertanegara Premium <i>Guest House</i> Menggunakan <i>Support Vector Machine</i>
Penulis	Wirdhayanti Paulina, Fitra Abdurrachman Bachtiar, Alfi Nur Rusydi

Judul Jurnal/ <i>Proceeding</i>	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer
Tahun Penerbitan	2020
Masalah Utama yang diangkat	Selama ini proses pengolahan ulasan pelanggan masih berfokus hanya pada <i>guest review</i> dan proses ini dinilai tidak efektif karena jumlah <i>guest review</i> yang sedikit dan informasi yang tersedia juga terbatas
Metode Ekstraksi	<i>preprocessing</i>
Metode Klasifikasi	<i>Support Vector Machine</i>
Hasil Penelitian dan Kesimpulan	<p>Hasil: Aspek layanan memiliki nilai akurasi tertinggi sebesar 0,83, sedangkan nilai akurasi terendah terdapat pada aspek harga sebesar 0,70. Aspek kamar memiliki nilai <i>precision</i> tertinggi pada sentimen positif sebesar 0,82, sedangkan nilai <i>precision</i> terendah pada sentimen positif terdapat pada aspek harga sebesar 0,65. Aspek layanan dan harga memiliki nilai <i>recall</i> tertinggi pada sentimen positif sebesar 0,92, sedangkan nilai <i>recall</i> terendah pada sentimen positif terdapat pada aspek kamar sebesar 0,75. Aspek layanan memiliki nilai <i>F1-Score</i> tertinggi pada sentimen positif sebesar 0,85, sedangkan nilai <i>F1-Score</i> terendah pada sentimen positif terdapat pada aspek harga sebesar 0,63.</p> <p>Kesimpulan: Pengujian hasil klasifikasi sentimen memiliki rata-rata yang baik dengan nilai <i>Accuracy</i>, <i>Precision</i>, <i>Recall</i>, dan <i>F1-Score</i> diatas 70%</p>