

# **Analisis Sentimen Berbasis Aspek Pada Twitter Menggunakan Metode Bidirectional Long Short-Term Memory**

**Tugas Akhir**

**diajukan untuk memenuhi salah satu syarat  
memperoleh gelar sarjana**

**dari Program Studi Sarjana Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**1301180279**

**Rizki Annas Sholehah**



**UNIVERSITAS  
Telkom**

**Program Studi Sarjana Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**Bandung**

**2023**

**LEMBAR PENGESAHAN**

**Analisis Sentimen Berbasis Aspek Pada Twitter Menggunakan Metode Bidirectional Long Short-Term Memory**

*Aspect-Based Sentiment Analysis on Twitter Using Bidirectional Long Short-Term Memory*

**1301180279**

**Rizki Annas Sholehat**

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada Program Studi Sarjana Informatika

Fakultas Informatika  
Universitas Telkom

Bandung, 24 Januari 2023

Menyetujui

Pembimbing I,

Pembimbing II,

**Dr. Erwin Budi Setiawan, M.T.**

**NIP: 00760045-1**

**Dr. Yuliant Sibaroni S.Si., M.T.**

**NIP: 00750036**

Ketua Program Studi  
Sarjana Informatika,

**Dr. Erwin Budi Setiawan, M.T.**

**NIP: 00760045-1**

**LEMBAR PERNYATAAN**

Dengan ini saya, Rizki Annas Sholehat, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Analisis Sentimen Berbasis Aspek Pada Twitter Menggunakan Metode *Bidirectional Long Short-Term Memory* beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 24 Januari 2023

Yang Menyatakan



Rizki Annas Sholehat

## **Analisis Sentimen Berbasis Aspek Pada Twitter Menggunakan Metode Bidirectional Long Short-Term Memory**

**Rizki Annas Sholehat<sup>1</sup>, Erwin Budi Setiawan<sup>2</sup>, Yuliant Sibaroni<sup>3</sup>**

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>rizkiannas@students.telkomuniversity.ac.id, <sup>2</sup>erwinbudisetiawan@telkomuniversity.ac.id,

<sup>3</sup>yuliant@telkomuniversity.ac.id

---





**Abstrak**

Twitter sebagai salah satu media sosial dengan user terbanyak di dunia seringkali digunakan sebagai media untuk berbagi opini yang dapat bersifat positif maupun negatif. Ulasan film yang mengandung banyak penjelasan kompleks dan penilaian akan sulit diklasifikasi. Maka dari itu diperlukan proses analisis sentimen berdasarkan aspek untuk menganalisis polaritas opini ulasan film berdasarkan aspek yang telah ditentukan. Penelitian ini dilakukan untuk menganalisis polaritas opini ulasan film berdasarkan aspek menggunakan metode Bidirectional Long Short-Term Memory dan ekstraksi fitur GloVe. Penelitian ini menggunakan aspek plot, akting dan direktor dengan total dataset sebanyak 17.247 data. Bidirectional Long Short-Term Memory terbukti menciptakan hasil yang relevan dan akurat untuk proses analisis sentimen dengan nilai akurasi terbesar sebesar 56,29% pada aspek plot, 87,07% pada aspek akting, dan 85,55% pada aspek direktor. Ditambah ekstraksi fitur dengan TF-IDF, ekspansi fitur dengan GloVe dan teknik Over-Sampling dengan SMOTE terbukti mampu meningkatkan nilai kinerja penelitian ini hingga 13,57% pada aspek plot, 4,16% pada aspek akting, dan 10,48% pada aspek direktor.

**Kata kunci : Deep learning; Bidirectional Long Short-Term Memory; GloVe; Analisis Sentimen; Aspek**

---



**Abstract**

Twitter as one of the social media with the most users in the world, is often used as a medium for sharing opinions that can be positive or negative. Movie reviews containing many complex explanations and judgments will be challenging to classify. Therefore a sentiment analysis process based on aspects is needed to analyze the polarity of film review opinions based on predetermined aspects. This research aims to analyze the polarity of film review opinions based on aspects using the Bidirectional Long Short-Term Memory method and GloVe feature extraction. This study uses plot, acting, and director aspects with a total dataset of 17.247 data. Bidirectional Long Short-Term Memory is proven to produce relevant and accurate results for sentiment analysis with the greatest accuracy of 56,29% in the plot aspect, 87,07% in the acting aspect, and 85,55% in the director aspect. GloVe feature extraction is proven to increase the performance value of this research by up to 13,57% in the plot aspect, 4,16% in the acting aspect, and 10,48% in the director aspect.

**Keywords:** Deep learning; Bidirectional Long Short-Term Memory; GloVe; Sentiment Analysis; Aspect

---



## 1. Pendahuluan

Di masa sekarang yang serba digital, masyarakat modern menggantungkan kebutuhan sehari-harinya pada teknologi. Media sosial adalah salah satu teknologi yang terus berkembang dan digunakan masyarakat luas. Pengguna media sosial dapat berkomunikasi, berbagi, bahkan menciptakan hal baru. Salah satu media sosial paling banyak digunakan saat ini adalah Twitter. Berdasarkan statistik yang dihimpun oleh [1], jumlah pengguna aktif media sosial Twitter di Indonesia menduduki peringkat keempat terbanyak di dunia. Ini membuktikan bahwa engagement pada pengguna aktif Twitter di Indonesia cukup tinggi.

Semakin mudahnya berbagi sesuatu pada media sosial, hal ini berdampak pula pada salah satu ranah dunia hiburan yang banyak digemari oleh sebagian besar kalangan masyarakat yaitu industri perfilman. Akses tersebut banyak dimanfaatkan pengguna Twitter untuk membuat sebuah tweet atau kicauan yang berisi ulasan tentang film yang telah mereka tonton sebelumnya. Ulasan tersebut dapat berupa ulasan yang bersifat positif maupun negatif. Semakin banyaknya informasi yang dibagikan oleh pengguna-pengguna Twitter dengan bahasa yang terkadang ambigu, maka akan sangat berpengaruh terhadap pemahaman konteks oleh pengguna lain ke sebuah film yang sedang diulas. Maka dari itu diperlukan analisis sentimen agar dapat mengklasifikasi ulasan film tersebut.

Analisis sentimen adalah sebuah proses mengesktraksi informasi tentang emosi atau perasaan dari seseorang dalam merespon suatu hal [2]. Kombinasi analisis sentimen saat ini telah sampai pada level penggabungan *entity* atau *aspect* di dalam penerapannya untuk mencapai hasil analisis yang lebih tepat. *Aspect-based sentiment analysis* (ABSA) memberikan kepastian hasil apakah dari sebuah opini yang positif juga menghasilkan opini positif secara keseluruhan aspek entitas di dalamnya [3]. Kombinasi ini mendukung penelitian ini karena ulasan film biasanya mengandung banyak aspek seperti genre, pemeran, alur cerita dan lain-lain.

Dalam pemrosesannya, saat ini analisis sentimen mulai menggunakan deep learning sebagai metode pemrosesan terbaru yang terus berkembang. Deep learning (DL) adalah metode machine learning yang mempelajari representasi data menggunakan algoritma artificial neural network yang dapat digunakan pada layer berlapis [4]. Model deep learning yang sering digunakan untuk analisis sentimen adalah Bidirectional Long Short-Term Memory (Bi-LSTM). Bi-LSTM adalah varian pengembangan dari model Long Short-Term Memory. Bi-LSTM memiliki 2 jenis input yaitu input forward dan input backward. Ini menjadikan Bi-LSTM dapat mempelajari informasi masa lampau (*past*) dan masa yang akan datang (*future*) dalam setiap sekuen input. Performa Bi-LSTM terbukti dari penelitian [5] yang dilakukan Cheng dan Tsai tentang analisis sentimen sosial media menggunakan beberapa model *deep learning* menunjukkan Bi-LSTM menghasilkan skor akurasi, presisi, *recall*, *specificity*, dan F1 tertinggi dibandingkan model deep learning lainnya.

Masalah yang diangkat pada tugas akhir ini adalah bagaimana pengaruh dan tingkat performansi setelah diterapkan teknik ekstraksi fitur TF-IDF dan ekspansi fitur dengan metode GloVe pada algoritma Bidirectional Long Short-Term Memory untuk menganalisis sentimen pada twitter yang berbasis aspek.

Batasan penelitian dalam Tugas Akhir ini adalah kumpulan data sentimen berbahasa Indonesia, total 17.247 tweet hanya berisi topik ulasan film, dan proses pelabelan sentimen dilakukan secara manual menjadi tiga kategori, yaitu positif, netral, dan negatif.

Tujuan dari penelitian ini adalah mengimplementasi, mengukur nilai performansi, serta menganalisis hasil sistem klasifikasi sentimen yang dibangun menggunakan metode ekstraksi fitur TF-IDF dan ekspansi fitur GloVe pada data sentimen Bahasa Indonesia dalam *dataset* tweet yang telah disiapkan.

## 2. Studi Terkait

Beberapa penelitian telah dilakukan sebelumnya terkait analisis sentimen dengan menggunakan *deep learning*. Proses klasifikasi ulasan tersebut akan mempermudah pengguna untuk mengkategorikan sebuah opini yang bersifat positif dan bersifat negatif secara lebih tepat.

Pada penelitian [6], Guixian Xu dkk melakukan penelitian analisis sentimen terhadap teks komentar situs penyedia layanan pemesanan hotel menggunakan metode Bi-LSTM. Penelitian tersebut menggunakan *Seninfo+TF-IDF* sebagai *word vector representation* terbaik secara performa pada model Bi-LSTM di proses sentimen analisis penelitian tersebut. Lalu, untuk membuktikan efektifitas dari metode Bi-LSTM, peneliti membandingkan performa Bi-LSTM dengan metode RNN, CNN, LSTM, dan Naive Bayesian. Metode Bi-LSTM memperoleh skor F1 92,18%, dibandingkan metode lain yang berada antara 84% hingga 89%.

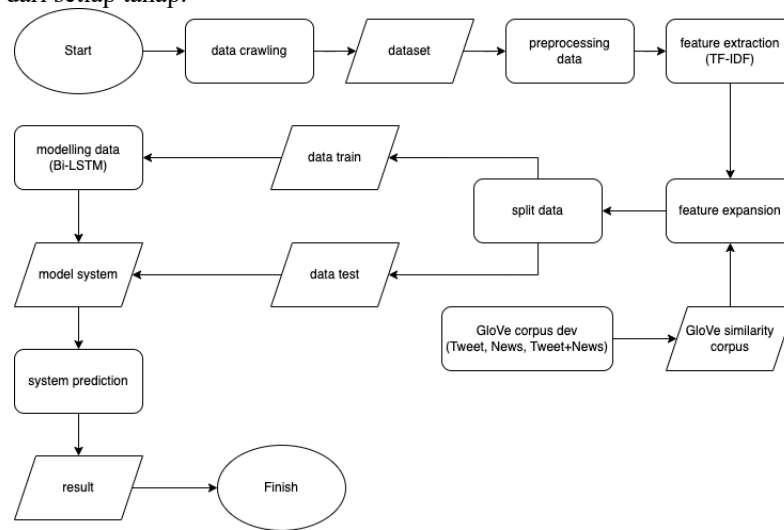
Penelitian lainnya, Kemal Hernandi dkk [7] melakukan penelitian analisis sentimen terhadap customer experience indihome pada Twitter menggunakan metode Bi-LSTM. Dalam implementasinya, peneliti membuat 2 model. Model 1 untuk mendeteksi sentimen negatif dan tidak negatif, dan model 2 untuk mendeteksi sentimen positif dan netral. Dari banyaknya konfigurasi model yang bisa dilakukan, RMSProp menghasilkan *optimizer* dan *learning rate* terbaik yaitu dengan *learning rate* 0,0001 dapat mencapai *test accuracy* sebesar 89,22% pada model 1 dan 88,2% pada model 2.

Penelitian selanjutnya terkait penggunaan Bi-LSTM dalam analisis sentimen [8], menyimpulkan bahwa Bi-LSTM lebih efektif dari CNN dan LSTM dalam mempelajari konteks dari kata demi kata di sebuah kalimat, karena metode ini menggabungkan *forward hidden layer* dan *backward hidden layer*.

Terkait dengan penggunaan aspek dalam analisis sentimen pada ulasan film, penelitian [9] memasukan aspek-aspek film seperti *screenplay*, *plot*, *movie*, *direction*, *acting*, dan *music*. Dihasilkan kesimpulan bahwa aspek terpenting pada penelitian tersebut adalah *plot*, *movie*, dan *acting*.

### 3. Sistem yang Dibangun

Sistem dibangun seperti pada Gambar 1. Sistem terbagi menjadi beberapa bagian, seperti *data crawling*, *data labelling*, *data preprocessing*, ekstraksi fitur, ekspansi fitur GloVe, split data dan pemodelan analisis sentimen. Berikut penjelasan dari setiap tahap.



Gambar 1 Sistem Analisis Sentimen

#### 3.1 Crawling dan Pelabelan Data

Proses crawling data dilakukan menggunakan Application Programming Interface (API) yang telah disediakan Twitter secara open-source melalui Twitter Developer. Kata kunci yang digunakan adalah hal-hal yang berkaitan dengan ketiga aspek yang telah ditentukan. Sebagai contoh untuk aspek plot berisi kata kunci "cerita" dan "alur", aspek akting berisi "aktor" dan "pemeran", aspek direktor berisi "sinematografi" dan "pembuatan film". Lalu data dilabeli dan divalidasi oleh 5 orang berdasarkan penilaian subjektif. Tabel 1 merupakan contoh data *tweet* yang telah dilabeli.

Tabel 1 Pelabelan Data

Tweet	Plot	Akting	Direktor
the pursuit of happyness verry nice y cerita yg menarik dan memotivasi	1	0	0
red notice jelek akting nya ga dapet semua	0	1	0

#### 3.2 Preprocessing Data

Di tahap ini data-data yang telah dikumpulkan akan disiapkan. Proses persiapan ini antara lain *data cleaning*, *data normalization*, *tokenization*, *case folding*, *stop word removal*, *stemming*.

- **Data Cleaning**  
Tanda baca, simbol, angka, emoji adalah bagian tweet yang akan mengurangi ketepatan hasil analisis data. Maka dari itu di proses ini akan dilakukan penghapusan hal-hal tersebut.
- **Data Normalization**  
Normalisasi data adalah tahap untuk mengidentifikasi kata yang terlalu banyak ejaannya yang tidak sesuai, yang kemudian akan diganti dengan ejaan yang sesuai KBBI.
- **Tokenization**  
Pada tahap ini tiap kata pada kalimat dipisahkan oleh spasi. Ini bertujuan untuk mendapatkan kata yang sering muncul pada suatu topik.
- **Case Folding**  
*Case folding* adalah mengubah seluruh huruf menjadi huruf kecil (lowercase).
- **Stop Word Removal**  
*Stop word removal* adalah proses menghilangkan kata-kata yang dianggap tidak penting. Kata-kata tidak penting ini tidak memiliki arti khusus.
- **Stemming**  
*Stemming* adalah tahapan mengubah kata yang memiliki imbuhan menjadi kata dasar.

### 3.3 Term Frequency - Inverse Document Frequency (TF-IDF)

Proses *extraction feature* merupakan faktor utama yang mempengaruhi keakuratan pada tahap klasifikasi. TF-IDF adalah fitur ekstraksi yang paling banyak digunakan saat ini, metode ini bekerja dengan cara menghitung bobot setiap kata yang umum digunakan [10]. Representasi suatu data tergantung pada jumlah fitur yang diperoleh di seluruh topik. Perhitungan TF-IDF ini didasarkan pada rumus berikut:

$$W_{ij} = t_{fij} \times Idf_j, Idf_j = \log\left(\frac{N}{df_j}\right) \quad (1)$$

### 3.4 Global Vector (GloVe)

Global Vector atau GloVe adalah model yang dapat menyimpan statistik kejadian munculnya kata secara global, dan model ini nantinya digunakan untuk merepresentasikan sebuah makna. Model GloVe dibangun karena pengamatan bahwa frekuensi kedekatan kata dapat mengarah pada sebuah kesimpulan. Nilai inilah yang digunakan untuk mengekspansi fitur sentimen analisis. Hasil dari algoritma GloVe ini akan menghasilkan output berupa list dari kata-kata yang dianggap memiliki *similarity*. Sebagai contoh, Tabel 2 adalah daftar kata-kata yang mirip dengan kata "film" dan telah peringkat.

**Tabel 2 Similarity Word**

Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9
cerita	dokumenter	sekuel	hostiles	keinget	animasi	mcu	superhero	sequel

### 3.5 Feature Expansion

Setelah membuat korpus yang dibuat dengan GloVe, pengertian pada fitur-fitur yang dimiliki dapat diperluas dengan menggunakan *feature expansion*. *Feature expansion* bekerja dengan mengganti vektor yang berisi nilai "0" dengan kata-kata yang dianggap memiliki kemiripan pada korpus GloVe tadi.

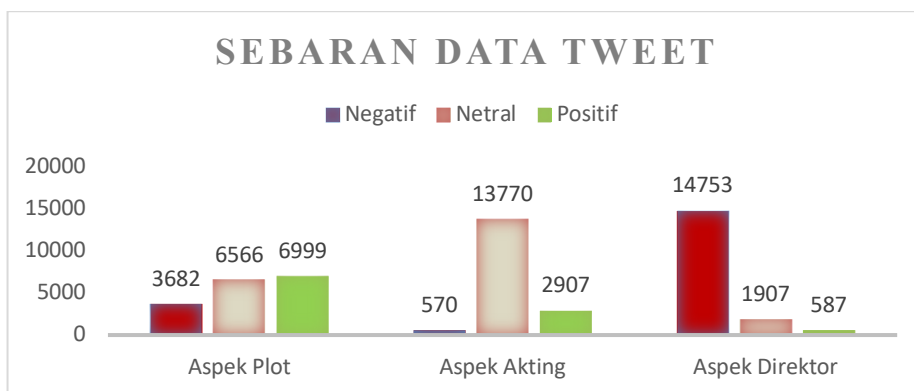
### 3.6 Bidirectional Long Short-Term Memory

*Bidirectional Long Short-Term Memory* (Bi-LSTM) merupakan varian turunan dari *Long Short Term Memory* (LSTM). Dibandingkan dengan LSTM, Bi-LSTM dapat menyelesaikan *task* pemodelan sekuensial lebih baik karena LSTM hanya mengeksplorasi konteks masa lalu atau *backward* saja [11]. Bi-LSTM dapat memasukan 2 input sekaligus ke dalam arsitektur, yaitu input *forward* dan *input backward*. Namun, dari arsitektur ini membuat *output* yang dikeluarkan hanya satu. Bi-LSTM memproses informasi sebelumnya dan informasi setelahnya secara dua arah. Proses selanjutnya adalah *feed forward neural* untuk pengklasifikasian yang lebih detail. Terdapat *hidden layer* yang saling berlawanan sehingga proses pelatihan akan lebih memahami data pada *time series*.

#### 4. Evaluasi

##### 4.1 Sebaran Data

Data tweet yang telah diperoleh dari proses data crawling berjumlah 17.274 tweet berbahasa Indonesia yang berfokus pada topik ulasan film. Gambar 2 menunjukkan hasil *labelling* dari data tweet.



Gambar 2 Sebaran Data

Lalu, data dalam pembuatan kamus kata menggunakan data yang diambil dari beberapa media berita dan data dari dataset tweet. Terdapat 3 Corpus GloVe yang dibuat untuk nantinya digunakan dalam pemodelan, diantaranya adalah corpus dengan dataset tweet, corpus dengan dataset berita, dan corpus dengan dataset tweet+berita.

##### 4.2 Hasil dan Analisis Pengujian

Di penelitian ini, terdapat 4 skenario yang akan dilakukan menggunakan klasifikasi Bidirectional Long Short-Term Memory. Skenario pertama akan membandingkan rasio antara 90:10, 80:20, dan 70:30 untuk split data test dan data train. Parameter perbandingan rasio ini adalah nilai akurasi dan F1 score terbaik. Skenario kedua dilakukan ekstraksi fitur menggunakan TF-IDF. Setelah dilakukan ekstraksi fitur, selanjutnya melakukan ekspansi fitur atau merubah angka 0 pada data agar dapat memiliki nilai similarity. Dan yang terakhir adalah menggabungkan teknik oversampling menggunakan SMOTE untuk melihat apakah masih ada kenaikan performansi jika dilakukan penyeimbangan data.

###### 4.2.1 Skenario Pertama (Baseline)

Skenario ini dilakukan untuk mencari rasio data train dan data test terbaik yang nantinya akan digunakan sebagai acuan rasio dalam menjalankan skenario berikutnya. Untuk aspek Plot dan Aktting menghasilkan nilai akurasi dan F1-score terbaik di rasio 90:10, sedangkan aspek Direktor menghasilkan nilai akurasi dan F1-score yang tidak jauh berbeda antar rasio namun nilai terbaik ada pada rasio 70:30.

Tabel 3 Skenario Pertama

Rasio	Aspek	Akurasi	F1-score
70:30	Plot	52,52%	45,82%
	Aktting	83,56%	42,85%
	Direktor	<b>85,55%</b>	<b>30,73%</b>
80:20	Plot	53,77%	42,74%
	Aktting	86,00%	48,89%
	Direktor	85,54%	30,73%
90:10	Plot	<b>56,29%</b>	<b>50,48%</b>
	Aktting	<b>87,07%</b>	<b>51,51%</b>

	Direktor	85,51%	30,72%
--	----------	--------	--------

#### 4.2.2 Skenario Kedua (TF-IDF)

Skenario selanjutnya adalah melakukan ekstraksi fitur menggunakan TF-IDF agar setiap kata memiliki bobot. Tabel 4 menunjukkan peningkatan pada akurasi dan F1-score yang cukup signifikan di seluruh aspek setelah TF-IDF diterapkan.

**Tabel 4 Skenario Kedua**

Aspek	Akurasi		F1-score	
	Sebelum	Sesudah	Sebelum	Sesudah
Plot	56,29%	<b>68,02%</b>	50,48%	<b>65,29%</b>
		(+11,73%)		(+14,81%)
Akting	87,07%	<b>89,32%</b>	51,51%	<b>65,29%</b>
		(+2,25%)		(+13,78%)
Direktor	85,55%	<b>88,79%</b>	30,73%	<b>51,71%</b>
		(+3,24%)		(+20,98%)

#### 4.2.3 Skenario Ketiga (Ekspansi Fitur GloVe)

Skenario ini bertujuan untuk memperluas kosakata dari suatu kata pada dataset. Dapat dilihat dari Tabel 5, Tabel 6, dan Tabel 7 untuk aspek plot dan akting mengalami peningkatan akurasi dan f1-score pada feature Top-1 dan untuk aspek direktor mengalami peningkatan pada feature Top-20. Peningkatan akurasi dan F1-score pada skenario ini tidak terlalu signifikan, peningkatan tertinggi hanya sebesar 2,36% pada aspek akting bagian F1-score.

**Tabel 5 Pengujian Ketiga Aspek Plot**

Feature	Akurasi				F1-score			
	Sebelum	Corpus Tweet	Corpus Berita	Corpus Tweet+Berita	Sebelum	Corpus Tweet	Corpus Berita	Corpus Tweet+Berita
Top-1	68,02%	66,75%	66,78%	<b>68,09%</b>	65,29%	63,85%	66,78%	<b>65,45%</b>
Top-5	68,02%	67,56%	67,66%	67,35%	65,29%	64,58%	67,66%	64,91%
Top-10	68,02%	67,27%	66,76%	67,55%	65,29%	64,65%	66,76%	64,96%
Top-20	68,02%	67,58%	67,65%	67,62%	65,29%	64,96%	67,65%	64,57%

**Tabel 6 Skenario Ketiga Aspek Akting**

Feature	Akurasi				F1-score			
	Sebelum	Corpus Tweet	Corpus Berita	Corpus Tweet+Berita	Sebelum	Corpus Tweet	Corpus Berita	Corpus Tweet+Berita
Top-1	89,32%	89,41%	89,47%	<b>89,68%</b>	63,20%	62,19%	64,82%	<b>65,56%</b>
Top-5	89,32%	89,11%	89,16%	89,43%	63,20%	63,77%	62,56%	62,38%
Top-10	89,32%	89,43%	89,34%	89,36%	63,20%	65,49%	64,06%	62,21%
Top-20	89,32%	89,26%	89,44%	89,02%	63,20%	63,95%	63,04%	62,21%

**Tabel 7 Skenario Ketiga Aspek Direktor**

Feature	Akurasi				F1-score			
	Sebelum	Corpus Tweet	Corpus Berita	Corpus Tweet+Berita	Sebelum	Corpus Tweet	Corpus Berita	Corpus Tweet+Berita
Top-1	88,79%	88,57%	88,60%	88,50%	51,71%	49,24%	51,19%	49,62%
Top-5	88,79%	88,51%	88,55%	88,66%	51,71%	50,38%	51,19%	50,32%
Top-10	88,79%	88,47%	88,51%	88,61%	51,71%	50,38%	50,19%	50,51%
Top-20	88,79%	88,65%	88,57%	<b>88,91%</b>	51,71%	51,39%	49,51%	<b>51,84%</b>
Top-30	88,79%	-	-	88,56%	51,71%	-	-	51,05%

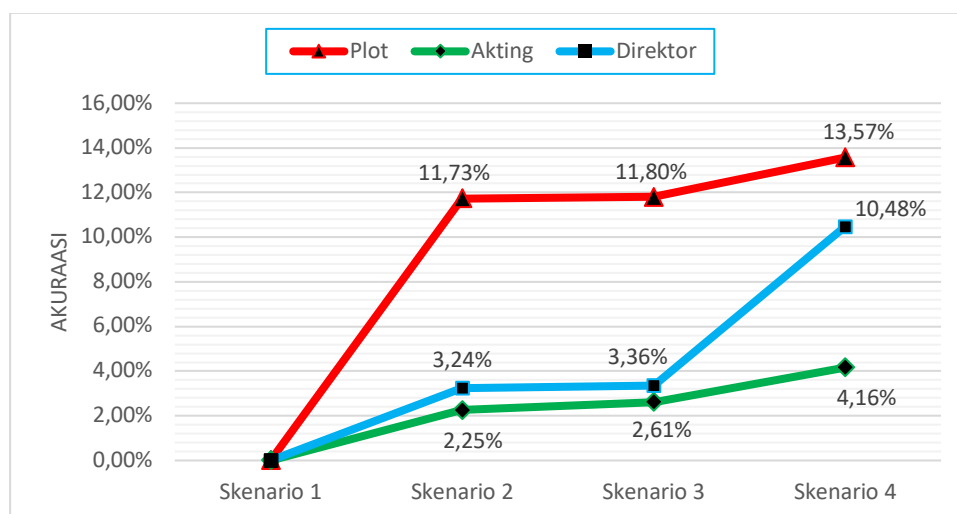
**4.2.4 Skenario Keempat (Oversampling SMOTE)**

Karena terdapat imbalance data pada dataset ini, maka skenario terakhir akan dilakukan teknik oversampling menggunakan SMOTE. Skenario ini dilakukan terakhir agar dapat melihat terlebih dahulu hasil dari TF-IDF dan GloVe secara murni terhadap dataset. Dapat terlihat di Tabel 8 seluruh aspek mengalami peningkatan signifikan pada akurasi dan F1-score.

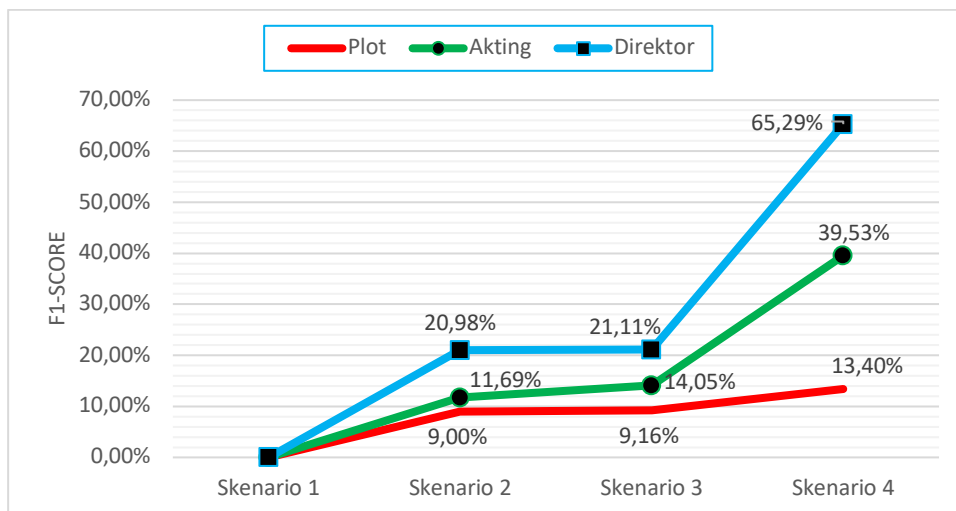
**Tabel 8 Skenario Keempat**

Aspek	Akurasi		F1-score	
	Sebelum	Sesudah	Sebelum	Sesudah
Plot	68,09%	<b>69,86%</b>	65,45%	<b>69,69%</b>
		(+1,77%)		(+4,24%)
Akting	89,68%	<b>91,23%</b>	65,56%	<b>91,04%</b>
		(+1,55%)		(+25,48%)
Direktor	88,91%	<b>96,03%</b>	51,84%	<b>96,02%</b>
		(+7,12%)		(+44,18%)

**4.2.5 Analisis Pengujian**



**Gambar 3 Data Peningkatan Akurasi**



**Gambar 4 Data Peningkatan F1-Score**

Pada skenario pertama, hasil pengujian pada data train dan data test dengan rasio 90:10, 80:20, dan 70:30 menghasilkan rasio 90:10 adalah yang terbaik untuk aspek Plot dan Akting, dan rasio 70:30 untuk aspek Direktor. Skenario kedua adalah menerapkan ekstraksi fitur TF-IDF, dan menghasilkan peningkatan yang sangat signifikan pada akurasi dan F1-score di seluruh aspek. Pada skenario ketiga, data diuji dengan menggunakan ekspansi fitur GloVe. Data diuji menggunakan 3 corpus yaitu corpus tweet, corpus berita, dan corpus tweet+berita, dalam skenario ini, hasil terbaik didapatkan dari corpus tweet+news pada fitur top-1 untuk aspek Plot dan Akting, lalu pada fitur top-20 untuk aspek Direktor. Nilai performansi di skenario ini mengalami peningkatan karena data dilatih terlebih dahulu dengan perluasan fitur-fitur dari corpus yang telah dibangun. Lalu skenario terakhir adalah pengujian data dengan teknik OverSampling menggunakan SMOTE. Di skenario ini performansi data yang dilatih nilainya meningkat secara signifikan, hal ini disebabkan oleh data yang diuji menjadi seimbang dan berpengaruh terhadap data yang diuji di skenario-skenario sebelumnya.



## 5. Kesimpulan

Setelah melewati empat skenario pengujian, bisa diambil konklusi bahwa setiap skenario pengujian bisa menaruh dampak terhadap kinerja berdasarkan contoh Bidirectional Long Short-Term Memory yg sudah dibangun. Penyeimbangan data menggunakan SMOTE bisa memberikan peningkatan nilai kinerja masing-masing aspek. Implementasi fitur perluasan menggunakan GloVe terbukti bisa meningkatkan accuracy (%) menjadi lebih tinggi. Pada aspek plot, nilai akurasi tertinggi yang dapat diraih adalah 69,86%, & nilai F1-score 69,69%, untuk aspek akting nilai akurasi tertinggi yang dapat diraih adalah 91,23% & nilai F1-score 91,04%, dan untuk aspek direktor nilai akurasi tertinggi yang dapat diraih adalah 96,03% & nilai F1-score 96,02%. Saran penelitian selanjutnya adalah mencoba menggunakan perpaduan metode ekstraksi fitur lainnya seperti Bag of Words (BOW) dan menggunakan ekspansi fitur lainnya seperti contohnya word2vec atau FastText. Sebagai catatan untuk membangun dataset agar lebih memperhatikan kualitas datanya agar menghasilkan akurasi yang lebih tinggi.

## DAFTAR PUSTAKA

- [1] H. Tankovska, "Twitter: most users by country | Statista," Jan. 2022. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> (accessed May 14, 2022).
- [2] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," *Expert Syst Appl*, vol. 118, pp. 272–299, Mar. 2019, doi: 10.1016/j.eswa.2018.10.003.
- [3] S. M. Jiménez-Zafra, M. T. Martín-Valdivia, E. Martínez-Cámara, and L. A. Ureña-López, "Combining resources to improve unsupervised sentiment analysis at aspect-level," *J Inf Sci*, vol. 42, no. 2, pp. 213–229, Apr. 2016, doi: 10.1177/0165551515593686.
- [4] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1253.
- [5] L.-C. Cheng and S.-L. Tsai, "Deep learning for automated sentiment analysis of social media," in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2019, pp. 1001–1004. doi: 10.1145/3341161.3344821.
- [6] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.
- [7] M. K. Hernandi, S. A. Wibowo, and S. Suyanto, "Sentiment Analysis Implementation For Detecting Negative Sentiment Towards Indihome In Twitter Using Bidirectional Long Short Term Memory," in *2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, Jul. 2021, pp. 143–147. doi: 10.1109/IAICT52856.2021.9532546.
- [8] H. Elfaik and E. H. Nfaoui, "Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 395–412, Dec. 2020, doi: 10.1515/jisys-2020-0021.
- [9] V. Parkhe and B. Biswas, "Aspect Based Sentiment Analysis of Movie Reviews: Finding the Polarity Directing Aspects," in *2014 International Conference on Soft Computing and Machine Intelligence*, Sep. 2014, pp. 28–32. doi: 10.1109/ISCMI.2014.16.
- [10] K. Kumar, B. S. Harish, and H. K. Darshan, "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, p. 109, 2019, doi: 10.9781/ijimai.2018.12.005.
- [11] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, Apr. 2019, doi: 10.1016/j.neucom.2019.01.078.