

## Ekstraksi Aspek menggunakan BiLSTM-CRFs pada Ulasan Lipstik Bahasa Indonesia

Rifqi Ahmad Fauzi<sup>1</sup>, Ade Romadhony<sup>2</sup>

<sup>1,2</sup> Universitas Telkom, Bandung

rifqiahmadfauzi@students.telkomuniversity.ac.id<sup>1</sup>, aderomadhony@telkomuniversity.ac.id<sup>2</sup>

### Abstrak

Ulasan merupakan suatu tulisan opini yang menilai mengenai karya atau kejadian yang diulas, dan aspek merupakan kata atau frasa yang menggambarkan atribut atau fitur entitas yang menjadi target opini. Di era modern, terdapat banyak situs yang memberi wadah terhadap pengguna untuk memberikan ulasan terhadap suatu produk, contohnya situs *femaledaily.com* yang memberikan wadah kepada penggunanya untuk mengulas produk-produk kecantikan, salah satunya produk lipstik. Aspek yang terdapat pada ulasan menjadi informasi berharga bagi para pemilik bisnis karena dengan adanya ulasan produk, mereka bisa mendapatkan wawasan dan informasi terhadap produknya sehingga dapat meningkatkan kualitas produknya. Namun mengambil informasi aspek dari teks ulasan dengan jumlah yang banyak membutuhkan waktu dan sumber daya yang besar jika dilakukan secara manual. Adanya sebuah sistem untuk dapat mengekstraksi aspek secara otomatis dari teks ulasan akan membantu. Oleh karena itu, pada penelitian ini dilakukan pembangunan sistem ekstraksi aspek dengan menggunakan metode *Bidirectional Long Short-Term Memory* dengan *Conditional Random Fields* (BiLSTM-CRFs). Hasil pengujian mendapatkan nilai *F1 Score* sebesar 57% dengan *tuning parameter* dimensi vektor *embedding layer* sebesar 256, banyak *batch* sebesar 32, dan banyak *epoch* sebesar 64.

**Kata kunci :** Ekstraksi aspek, *Conditional Random Field*, Ulasan Lipstik, Bahasa Indonesia, *Bidirectional Long Short-Term Memory*, *BiLSTM-CRFs*

### Abstract

A review is an opinion piece that evaluates the work or event being reviewed. An aspect is a word or phrase that describes the entity's attributes or features that is the opinion's target. In the modern era, many sites provide a forum for users to review a product; for example, the *femaledaily.com* site provides a forum for users to review beauty products, lipstick. Aspects contained in reviews are valuable information for business owners because, with product reviews, they can get insight and information about their products to improve the quality of their products. However, extracting aspect information from a large amount of review text requires a significant amount of time and resources if done manually. Having a system to be able to extract aspects automatically from the review text would be helpful. Therefore, this study developed an aspect extraction system using the *Bidirectional Long Short-Term Memory* method with *Conditional Random Fields* (BiLSTM-CRFs). The test results get an *F1 Score* value of 57% with tuning of the vector embedding layer dimension parameters of 256, the number of batches of 32, and the number of epochs of 64.

**Keywords:** Aspect term extraction, *Conditional Random Field*, Lipstick Review, Indonesian language, *Bidirectional Long Short-Term Memory*, *BiLSTM-CRFs*

## 1. Pendahuluan

### Latar Belakang

Ulasan merupakan sebuah teks yang berfungsi untuk mengukur, menilai, dan memikirkan kritik mengenai karya atau kejadian yang diulas [1]. Sebuah teks ulasan umumnya mengandung informasi opini terhadap aspek produk. Aspek merupakan kata atau frasa yang menggambarkan atribut atau fitur entitas yang menjadi target opini. Opini merupakan kata atau frasa yang menunjukkan emosi subjektif terhadap atribut atau fitur entitas [3].

Sebuah ulasan produk berguna bagi konsumen untuk menentukan apakah produk tersebut layak untuk dibeli atau tidak. Menurut survei yang dilakukan *Bright Local*, hanya 48% konsumen yang akan melanjutkan transaksi jika penilaian produk kurang dari empat dalam skala lima [2]. Hal ini penting bagi pemilik bisnis karena dengan adanya ulasan produk, mereka mendapatkan wawasan dan informasi terhadap produknya sehingga dapat meningkatkan kualitas produknya. Namun dengan banyaknya ulasan yang terdapat pada situs-situs ulasan seperti *femaledaily.com* membuat pemilik bisnis kesulitan untuk memutuskan aspek apa saja yang sering diulas dalam produknya membutuhkan waktu dan sumber daya yang besar jika dilakukan secara manual. Dengan adanya masalah tersebut, akan dibangun sebuah sistem yang menangani ekstraksi aspek pada ulasan produk lipstik.

Ekstraksi aspek merupakan subtugas dari analisis sentimen yang terdiri dari pengidentifikasian target opini dalam teks opini [4]. Terdapat berbagai metode yang telah diterapkan untuk melakukan ekstraksi aspek, salah

satunya menggunakan *Pattern Knowledge* dan *Taxonomy* yang menghasilkan *F1 score* sekitar 20%-69%, dan *Recall* sebesar 89,3% [5].

Dalam Tugas Akhir ini, ekstraksi aspek dilakukan dengan metode *Bidirectional Long Short-Term Memory* dengan *Conditional Random Fields* (BiLSTM-CRFs). Dalam penelitian yang dilakukan oleh Athanasios [7], BiLSTM-CRFs memberikan performansi yang tinggi dalam melakukan ekstraksi aspek pada dataset ulasan laptop dan restoran yang menghasilkan *F1 Score* sebesar 88%. BiLSTM dirancang untuk menyandikan setiap elemen dengan mempertimbangkan konteks kiri dan kanan. CRFs memiliki keunggulan dibandingkan metode lainnya seperti *Hidden Markov Model* dalam melakukan segmentasi, dan pelabelan secara terurut [6].

### Topik dan Batasannya

Berdasarkan dari latar belakang tersebut, dapat dirumuskan permasalahan yang dapat diangkat dalam tugas akhir ini, yaitu bagaimana membangun dan mengimplementasikan sistem yang dapat mengekstraksi aspek menggunakan metode *Bidirectional Long Short-Term Memory* dengan *Conditional Random Fields* (BiLSTM-CRFs). Adapun Batasan-batasan yang ada dari permasalahan ini adalah

- Dataset yang digunakan merupakan ulasan produk lipstick Bahasa Indonesia dari situs *femaledaily.com*.
- Dataset terdiri dari 1.429 kalimat, lalu ditokenisasi menjadi 23.004 token/kata.
- Pelabelan aspek menggunakan notasi *beginning-inside-outside* (BIO) dan dibagi kedalam enam aspek, yaitu warna, kelembaban, pigmentasi, kemasan, harga, dan ketahanan warna. Tabel 1 menampilkan informasi label setiap aspek.

**Tabel 1. Informasi Label**

Label	Aspek
B-ASP-COL	Warna
I-ASP-COL	
B-ASP-MOI	Kelembaban
I-ASP-MOI	
B-ASP-PGM	Pigmentasi
I-ASP-PGM	
B-ASP-PKG	Kemasan
I-ASP-PKG	
B-ASP-PRI	Harga
I-ASP-PRI	
B-ASP-STA	Ketahanan warna
I-ASP-STA	
O	Tidak ada aspek

### Tujuan

Tujuan dari penelitian ini adalah untuk membangun sebuah sistem yang dapat mengekstraksi aspek menggunakan metode *Bidirectional Long-Short Term Memory* dengan *Conditional Random Fields* (BiLSTM-CRFs) berdasarkan hasil kinerja *F1 Score* dan melakukan analisis terhadap hasil prediksi.

### Organisasi Tulisan

Laporan tugas akhir ini terdiri dari lima bagian. Pada bagian pertama menjelaskan pendahuluan, pada bagian kedua menjelaskan studi terkait, pada bagian ketiga menjelaskan sistem yang dibangun, pada bagian keempat menjelaskan evaluasi dari sistem yang dibangun, dan pada bagian kelima menjelaskan kesimpulan yang diambil dari tugas akhir ini.

## 2. Studi Terkait Ekstraksi Aspek

Penelitian yang dilakukan oleh Tubisat [9] memberikan gambaran tentang teknik ekstraksi aspek implisit dari perspektif yang berbeda. Perspektif pertama adalah membuat analisis komparatif untuk teknik yang tersedia dalam mengekstraksi istilah implisit dengan ringkasan singkat masing-masing teknik. Perspektif kedua adalah untuk mengklasifikasikan dan membandingkan kinerja, dataset, bahasa yang digunakan, dan kekurangan teknik yang tersedia.

Ekstraksi aspek pada ulasan produk menggunakan metode *rule-based* yang dilakukan oleh Soujanya Poria [11] memberikan hasil bahwa metode *rule-based* dapat melakukan ekstraksi pada aspek eksplisit dan implisit.

Peneliti mengidentifikasi *implicit aspect clue* (IAC) pada setiap ulasan yang selanjutnya digabungkan dengan aspek terkait. Dengan hasil evaluasi model sebesar 92% untuk *precision* dan 94% untuk *recall*.

Ekstraksi aspek dapat dilakukan dengan menggunakan model *semi-supervised*, seperti yang telah dilakukan oleh Arjun Mukherjee [12], penelitian tersebut menggunakan model *Seeded Aspect and Sentiment* (SAS) dan *Maximum Entropy Seeded Aspect and Sentiment* (ME-SAS) pada dataset ulasan hotel *tripadvisor.com* yang terdiri dari 101.234 ulasan dan 692.783 kalimat. Penelitian ini juga menunjukkan bahwa model ME-SAS tidak perlu bantuan dari pengguna dalam tahapan pelatihan *maximum-entropy*.

### BiLSTM-CRFs

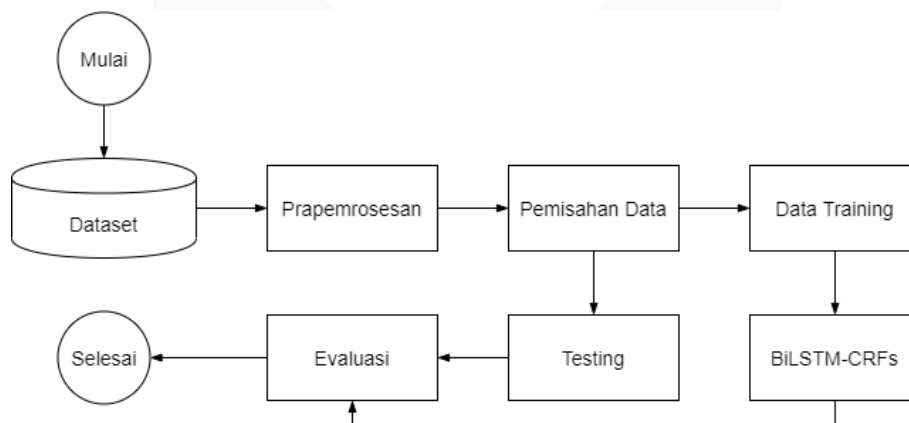
Dalam penelitian yang dilakukan oleh Athanasios [7], BiLSTM-CRFs memberikan performansi yang tinggi dalam melakukan ekstraksi aspek pada dataset ulasan laptop dan restoran yang menghasilkan F1 Score sebesar 88%. Dalam penelitian tersebut dataset dilabeli secara otomatis dengan kamus sentimen yang telah dibangun.

BiLSTM-CRFs dapat digunakan untuk meningkatkan analisis sentimen melalui klasifikasi tipe kalimat. Dalam penelitian ini [8] dilakukan sebuah teknik *divide-and-conquer* yang dapat mengklasifikasikan kalimat menjadi tiga tipe berdasarkan jumlah kemunculan target sentimen pada kalimat tersebut.

Dalam penelitian yang dilakukan oleh Miao [10], dilakukan sebuah analisis sentimen berbasis aspek menggunakan BiLSTM-CRFs pada 6357 ulasan ponsel. Pada penelitian ini menggabungkan vektor karakter Cina dan fitur posisi kata Cina yang mana dapat mengekstraksi kata atribut dan kata sentimen secara bersamaan, selagi mengekstraksi polarisasi dari kata sentimen. Sistem yang dibangun pada penelitian ini mendapatkan hasil evaluasi model sebesar 87,92% untuk *precision*, 88,05% untuk *recall*, dan 79,9% untuk *F1 Score*.

### 3. Sistem yang Dibangun

Sistem ekstraksi aspek ini dibangun menggunakan metode *Bidirectional Long-Short Term Memory* dengan *Conditional Random Fields* (BiLSTM-CRFs). Alur penelitian ini diilustrasikan pada Gambar 1.



**Gambar 1. Sistem pada ekstraksi aspek pada ulasan produk lipstik berbahasa Indonesia menggunakan BiLSTM-CRFs**

#### Dataset

Dataset merupakan ulasan produk lipstik dari situs *femaledaily.com* dengan jumlah ulasan sebanyak 250 ulasan, ulasan tersebut kemudian dipecah menjadi 1.429 kalimat, lalu ditokenisasi menjadi 23.004 token/kata. Setiap token tersebut kemudian dilabeli menggunakan *Named-entity recognition* (NER) dengan pendekatan notasi BIO seperti yang ditampilkan pada tabel 1 diatas. Berikut contoh bentuk isi dataset yang ditampilkan pada tabel 2 dibawah ini.

**Tabel 2. Contoh isi dataset**

Sentence_idx	word	tag
0	selalu	O
0	perlu	O
0	pinkish	B-ASP-COL
0	nude	I-ASP-COL
0	lipstick	O



*Integer* pada setiap kalimat merepresentasikan pemetaan kata yang bersangkutan dengan *index* kata tersebut. Jika kata dalam kalimat tersebut tidak sebanyak dengan panjang *padding*, maka akan diisi dengan '<OOV>', dengan *index* paling akhir, yaitu 2.913.

Pemisahan data

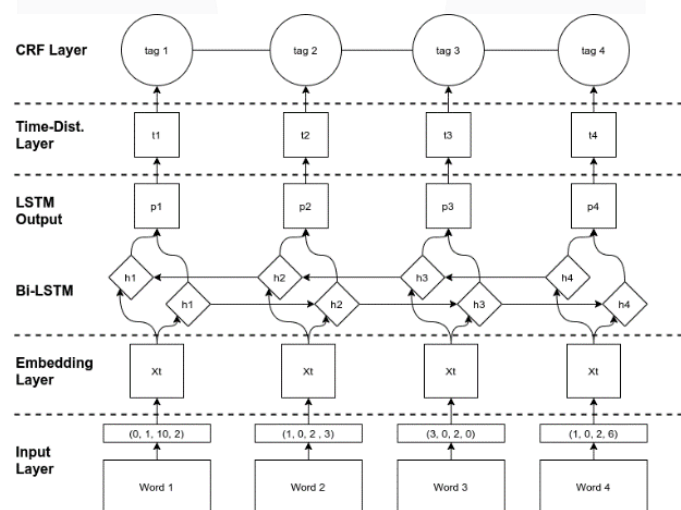
Setelah dataset melalui tahap prapemrosesan, maka dataset akan dipisahkan menjadi data latih dan data uji dengan perbandingan 75:25, dengan data latih sebanyak 1.071 kalimat, dan data uji sebanyak 358 kalimat. Persebaran label untuk data latih dan data uji ditampilkan pada tabel 6 dibawah ini.

**Tabel 6. Persebaran label berdasarkan data latih dan data uji**

Label	Variasi Label	Data latih	Data uji
B-ASP-COL	116	697	238
I-ASP-COL	263	675	230
B-ASP-MOI	72	267	55
I-ASP-MOI	70	207	54
B-ASP-STA	40	197	93
I-ASP-STA	48	176	93
B-ASP-PKG	25	102	32
I-ASP-PKG	58	65	28
B-ASP-PGM	19	99	35
I-ASP-PGM	17	48	15
B-ASP-PRI	17	97	53
I-ASP-PRI	23	39	27

BiLSTM-CRFs

Pada tahap ini akan dilakukan pelatihan pada data latih. BiLSTM-CRFs dapat secara efisien menggunakan *input feature* melalui BiLSTM dan informasi tag kalimat melalui CRFs. Berikut merupakan arsitektur yang dibangun pada penelitian ini.



**Gambar 2. Arsitektur BiLSTM-CRFs**

*Input layer* merupakan *layer* pertama untuk membaca sebuah *input*. Dalam penelitian ini, *input layer* akan menerima sebuah vektor *integer*, yaitu dataset yang telah melalui tahap prapemrosesan dengan panjang *padding* yang telah ditentukan, yaitu sebesar 80.

*Embedding layer* merupakan *hidden layer* yang menerima tiga argumen, yang pertama yaitu *input dimension* yang menerima sebuah nilai dari banyaknya variasi token/kata dalam dataset, yaitu sebesar 2914, yang kedua yaitu *output dimension* yang menerima sebuah nilai sebagai sebuah vektor dari setiap token/kata yang mana akan menjadi nilai *output* dari *layer* ini, dan yang ketiga yaitu *input length* yang menerima nilai dari panjang sekuen dataset, yaitu sebesar 86. *Embedding layer* akan mentransformasikan setiap token menjadi vektor dengan panjang yang telah ditentukan.

*Bidirectional LSTM layer* menerima *layer* LSTM pertama sebagai argumen, *layer* ini menerima *input* dari *output embedding layer* sebelumnya. Sebuah *bidirectional LSTM* mempunyai *forward* dan *backward output*, maka *output* tersebut akan digabungkan (*concatenate*) sebelum masuk kedalam *layer* selanjutnya.

*Time distributed layer* memungkinkan sebuah *output* dari setiap urutan *input* dari *layer* sebelumnya, yaitu *layer* BiLSTM, sehingga nilai dari setiap urutan *input* (*timestep*) dapat memprediksi sebuah sekuen yang diberikan sebagai *input*. Sehingga *output* memiliki label sebanyak urutan *input* [13].

CRFs *layer* mempunyai performansi yang tinggi dalam pelabelan secara terurut, penggunaan *layer* CRFs dapat membantu *layer* BiLSTM dalam proses klasifikasi.

#### Evaluasi

Penelitian ini merupakan salah satu bentuk klasifikasi, sehingga hasil dari proses klasifikasi harus dievaluasi. Cara yang biasa dilakukan untuk mengevaluasi dengan evaluasi matriks. Hasil dari klasifikasi dibagi menjadi empat jenis, yaitu *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN). Keempat jenis hasil klasifikasi digunakan untuk menentukan *accuracy*, *precision*, *recall*, dan *F1 Score* pada persamaan berikut.

$$Accuracy = \frac{TP + FN}{TP + FP + FN + TN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Berdasarkan persamaan (2), *precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dalam penelitian ini, *precision* digunakan untuk mengukur seberapa banyak suatu token yang tidak tepat dilabeli terhadap label tertentu (*false positive*). Berdasarkan persamaan (3), *recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Dalam penelitian ini, *recall* digunakan untuk mengukur seberapa banyak token dengan label tertentu yang tidak tepat dilabeli (*false negative*). *F1 Score* digunakan untuk mendapatkan nilai *harmonic mean* antara *precision* dengan *recall*, secara representasi, semakin tinggi *F1 score* maka semakin baik juga *precision* dan *recall* dari model yang dibangun.

#### 4. Evaluasi

Dalam penelitian ini, dilakukan pengujian sistem yang digunakan untuk mengetahui tingkat keberhasilan terhadap sistem yang telah dibangun ditinjau dari perolehan rata-rata *F1 Score* dari setiap label. Terdapat delapan skenario yang dilakukan pada penelitian ini untuk mendapatkan hasil terbaik. Skenario yang dilakukan adalah *tuning* jumlah vektor pada *embedding layer*, jumlah *epoch*, dan jumlah *batch* pada arsitektur yang dibangun. Tabel 6 dibawah ini menampilkan informasi *tuning* yang dilaksanakan pada setiap skenario pengujian di penelitian ini.

Setiap tabel hasil pengujian skenario terdapat kolom *precision* yang menampilkan nilai performansi sistem berdasarkan banyaknya *false positive* pada label yang terkait, kolom *recall* yang menampilkan nilai performansi sistem berdasarkan banyaknya *false negative* pada label yang terkait, kolom *f1 score* sebagai nilai *harmonic mean* antara *precision* dan *recall*, dan kolom *support* yang menampilkan banyaknya label pada data uji.

**Tabel 6. Informasi tuning pada setiap skenario pengujian**

Skenario	Vektor embedding layer	Batch	Epoch
1	256	16	32
2	256	16	64
3	256	32	32
4	256	32	64
5	384	16	32
6	384	16	64
7	384	32	32
8	384	32	64

##### 4.1 Pengujian Skenario 1

**Tabel 7. Hasil Pengujian Skenario 1**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.62	0.73	0.67	238
B-ASP-MOI	0.23	0.56	0.32	55
B-ASP-PGM	0.45	0.29	0.35	35

B-ASP-PKG	0.83	0.62	0.71	32
B-ASP-PRI	0.56	0.50	0.53	53
B-ASP-STA	0.75	0.49	0.60	93
I-ASP-COL	0.50	0.30	0.38	230
I-ASP-MOI	0.39	0.44	0.41	54
I-ASP-PGM	0.45	0.33	0.38	15
I-ASP-PKG	0.00	0.00	0.00	28
I-ASP-PRI	1.00	0.12	0.21	27
I-ASP-STA	0.93	0.57	0.71	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 58%, recall sebesar 57%, dan F1-Score sebesar 56%.

#### 4.2 Pengujian Skenario 2

**Tabel 8. Hasil Pengujian Skenario 2**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.60	0.71	0.65	238
B-ASP-MOI	0.27	0.60	0.37	55
B-ASP-PGM	0.72	0.80	0.76	35
B-ASP-PKG	0.86	0.78	0.82	32
B-ASP-PRI	0.79	0.78	0.79	53
B-ASP-STA	0.69	0.59	0.64	93
I-ASP-COL	0.51	0.37	0.43	230
I-ASP-MOI	0.34	0.48	0.40	54
I-ASP-PGM	0.36	0.60	0.45	15
I-ASP-PKG	0.25	0.04	0.06	28
I-ASP-PRI	0.57	0.32	0.41	27
I-ASP-STA	0.79	0.61	0.69	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 58%, recall sebesar 57%, dan F1-Score sebesar 56%.

#### 4.3 Pengujian Skenario 3

**Tabel 9. Hasil Pengujian Skenario 3**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.60	0.77	0.68	238
B-ASP-MOI	0.32	0.45	0.38	55
B-ASP-PGM	0.73	0.69	0.71	35
B-ASP-PKG	0.00	0.00	0.00	32
B-ASP-PRI	1.00	0.02	0.04	53
B-ASP-STA	0.67	0.55	0.60	93
I-ASP-COL	0.47	0.34	0.39	230
I-ASP-MOI	0.34	0.46	0.39	54
I-ASP-PGM	1.00	0.27	0.42	15
I-ASP-PKG	0.00	0.00	0.00	28
I-ASP-PRI	0.00	0.00	0.00	27
I-ASP-STA	0.81	0.56	0.66	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 55%, recall sebesar 47%, dan F1-Score sebesar 47%.

#### 4.4 Pengujian Skenario 4

**Tabel 10. Hasil Pengujian Skenario 4**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.60	0.76	0.67	238
B-ASP-MOI	0.24	0.64	0.34	55
B-ASP-PGM	0.79	0.63	0.70	35
B-ASP-PKG	0.80	0.75	0.77	32
B-ASP-PRI	0.88	0.78	0.82	53
B-ASP-STA	0.67	0.66	0.66	93
I-ASP-COL	0.56	0.38	0.45	230

I-ASP-MOI	0.31	0.50	0.38	54
I-ASP-PGM	0.33	0.53	0.41	15
I-ASP-PKG	0.08	0.04	0.05	28
I-ASP-PRI	0.78	0.28	0.41	27
I-ASP-STA	0.90	0.66	0.76	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 60%, recall sebesar 58%, dan F1-Score sebesar 57%.

#### 4.5 Pengujian Skenario 5

**Tabel 11. Hasil Pengujian Skenario 5**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.59	0.70	0.64	238
B-ASP-MOI	0.30	0.58	0.40	55
B-ASP-PGM	0.82	0.51	0.63	35
B-ASP-PKG	0.79	0.59	0.68	32
B-ASP-PRI	0.88	0.65	0.74	53
B-ASP-STA	0.70	0.61	0.66	93
I-ASP-COL	0.50	0.30	0.37	230
I-ASP-MOI	0.38	0.41	0.39	54
I-ASP-PGM	0.44	0.47	0.45	15
I-ASP-PKG	0.00	0.00	0.00	28
I-ASP-PRI	1.00	0.16	0.28	27
I-ASP-STA	0.86	0.59	0.70	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 60%, recall sebesar 51%, dan F1-Score sebesar 53%.

#### 4.6 Pengujian Skenario 6

**Tabel 12. Hasil Pengujian Skenario 6**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.64	0.74	0.69	238
B-ASP-MOI	0.28	0.58	0.38	55
B-ASP-PGM	0.70	0.80	0.75	35
B-ASP-PKG	0.85	0.72	0.78	32
B-ASP-PRI	0.82	0.67	0.73	53
B-ASP-STA	0.70	0.59	0.64	93
I-ASP-COL	0.56	0.31	0.40	230
I-ASP-MOI	0.36	0.37	0.36	54
I-ASP-PGM	0.39	0.47	0.42	15
I-ASP-PKG	0.50	0.04	0.07	28
I-ASP-PRI	0.60	0.24	0.34	27
I-ASP-STA	0.77	0.58	0.66	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 61%, recall sebesar 54%, dan F1-Score sebesar 55%.

#### 4.7 Pengujian Skenario 7

**Tabel 13. Hasil Pengujian Skenario 7**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.68	0.73	0.70	238
B-ASP-MOI	0.20	0.45	0.28	55
B-ASP-PGM	0.71	0.69	0.70	35
B-ASP-PKG	0.92	0.38	0.53	32
B-ASP-PRI	0.81	0.63	0.71	53
B-ASP-STA	0.72	0.46	0.56	93
I-ASP-COL	0.48	0.28	0.26	230
I-ASP-MOI	0.34	0.43	0.28	54
I-ASP-PGM	0.00	0.00	0.00	15
I-ASP-PKG	0.00	0.00	0.00	28
I-ASP-PRI	0.67	0.16	0.26	27



I-ASP-STA	0.80	0.40	0.53	93
-----------	------	------	------	----

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 59%, recall sebesar 46%, dan F1-Score sebesar 50%.

#### 4.8 Pengujian Skenario 8

**Tabel 14. Hasil Pengujian Skenario 8**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.58	0.76	0.65	238
B-ASP-MOI	0.23	0.55	0.33	55
B-ASP-PGM	0.77	0.69	0.73	35
B-ASP-PKG	0.65	0.69	0.67	32
B-ASP-PRI	0.85	0.65	0.74	53
B-ASP-STA	0.64	0.59	0.60	93
I-ASP-COL	0.46	0.34	0.39	230
I-ASP-MOI	0.27	0.46	0.34	54
I-ASP-PGM	0.28	0.60	0.46	15
I-ASP-PKG	0.12	0.04	0.06	28
I-ASP-PRI	0.67	0.16	0.26	27
I-ASP-STA	0.78	0.63	0.70	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 55%, recall sebesar 57%, dan F1-Score sebesar 53%.

#### 4.9 Perbandingan dengan LSTM-CRFs

Dibuat sebuah sistem arsitektur LSTM-CRFs dengan dataset dan skenario tuning yang sama seperti sistem yang dibangun, tabel 15 dibawah menampilkan hasil terbaik yang diperoleh oleh arsitektur LSTM-CRFs.

**Tabel 15. Hasil Terbaik LSTM-CRFs**

Label	Precision	Recall	F1-Score	Support
B-ASP-COL	0.68	0.84	0.75	238
B-ASP-MOI	0.24	0.42	0.30	55
B-ASP-PGM	0.65	0.63	0.64	35
B-ASP-PKG	0.63	0.75	0.69	32
B-ASP-PRI	0.92	0.66	0.77	53
B-ASP-STA	0.72	0.60	0.65	93
I-ASP-COL	0.56	0.40	0.47	230
I-ASP-MOI	0.47	0.31	0.38	54
I-ASP-PGM	0.00	0.00	0.00	15
I-ASP-PKG	0.00	0.00	0.00	28
I-ASP-PRI	0.00	0.00	0.00	27
I-ASP-STA	0.79	0.63	0.70	93

Dengan perolehan diatas maka rata-rata skor untuk precision sebesar 59%, recall sebesar 56%, dan F1-Score sebesar 56%. Kinerja dari model LSTM-CRFs tersebut berada dibawah kinerja model BiLSTM-CRFs skenario 4 yang ditampilkan pada tabel 10.

#### 4.10 Hasil output dari sistem yang dibangun

Berikut merupakan hasil *output* dari sistem yang dibangun terhadap data uji, tabel 16 menampilkan beberapa contoh data uji dengan label aktual, dan label dari sistem yang dibangun.

**Tabel 16. Contoh hasil labeling dari sistem yang dibangun**

Kalimat	Label Aktual	Label Prediksi Sistem
jade, ini, kalau, di, bibir, saya, warnanya, jadi, pink, terang	O, O, O, O, O, O, B-ASP-COL, I-ASP-COL, I-ASP-COL, I-ASP-COL	O, O, O, O, O, O, B-ASP-COL, O, B-ASP-COL, I-ASP-COL
kalau, lagi, bosan, sama, lipstick, yang, warnawarna, barbie, atau, warna, biasa, shades, copenhagen, ini, jadi, pilihannya	O, O, O, O, O, O, O, O, O, B-ASP-COL, I-ASP-COL, B-ASP-COL, I-ASP-COL, I-ASP-COL, O, O	O, O, O, O, O, O, B-ASP-COL, O, O, B-ASP-COL, O, O, I-ASP-COL, O, O, O
lipstick, kesukaan, karena, murah, dan, juga, warnanya, oke	O, O, O, B-ASP-PRI, O, O, B-ASP-COL, O	O, O, O, B-ASP-PRI, O, O, B-ASP-COL, O



pada data uji dibandingkan dengan *support* label tersebut. Berdasarkan data dari tabel 16 diatas, label I-ASP-PKG, I-ASP-PRI, dan I-ASP-COL merupakan label-label dengan variasi token yang hanya terdapat pada data uji paling tinggi.

**Tabel 18. Variasi token yang memiliki lebih dari satu label (multi label)**

Label	Token	Error Rate
B-ASP-COL	'agak', 'antwerp', 'athens', 'beige', 'berwarna', 'blue', 'bold', 'bright', 'brownish', 'cerah', 'coklat', 'color', 'colour', 'coral', 'dark', 'dead', 'deep', 'dusty', 'fuschia', 'fushia', 'harganya', 'hint', 'hitam', 'ini', 'light', 'lip', 'lipstick', 'liquid', 'matte', 'mauve', 'merah', 'merahnya', 'muda', 'muted', 'natural', 'ngejreng', 'nude', 'opaque', 'orange', 'orangnya', 'oranye', 'peach', 'peachy', 'perfect', 'pink', 'pinkish', 'pinknya', 'plum', 'pucat', 'red', 'ruby', 'russian', 'sama', 'satin', 'sedikit', 'semua', 'shade', 'sleek', 'stockholm', 'super', 'tahan', 'terlalu', 'terlihat', 'tidak', 'tua', 'ungu', 'velvet', 'warm', 'warna', 'warnanya'	58,62%
B-ASP-MOI	'agak', 'bibir', 'bold', 'buttery', 'creamy', 'creme', 'cukup', 'dead', 'deep', 'dry', 'extra', 'finish', 'finishing', 'formula', 'glossy', 'hasilnya', 'kering', 'lembab', 'lembut', 'lip', 'lipstick', 'lipstik', 'liquid', 'matte', 'mattenya', 'melembabkan', 'moist', 'moisturizing', 'nude', 'sangat', 'semi', 'semimatte', 'sleek', 'smooth', 'soft', 'super', 'tekstur', 'texture', 'very', 'walaupun'	52,78%
B-ASP-PGM	'cukup', 'matte', 'pigmentasi', 'pigmented', 'sangat', 'sekali', 'super', 'tekstur', 'very', 'warna', 'warnanya'	57,89%
B-ASP-PKG	'aplikatornya', 'bentuknya', 'kemasannya', 'lipstick', 'long', 'packaging', 'packagingnya', 'pecah'	32%
B-ASP-PRI	'affordable', 'harga', 'harganya', 'lipstick', 'mahal', 'murah', 'price', 'produk', 'segi'	52,94%
B-ASP-STA	'awet', 'awetnya', 'cukup', 'hilang', 'kering', 'ketahanan', 'kurang', 'lasting', 'lebih', 'long', 'longlasting', 'lumayan', 'nempel', 'pudar', 'segi', 'stay', 'staying', 'super', 'tahan', 'tetep', 'tidak', 'very', 'warna', 'warnanya'	60%
I-ASP-COL	'agak', 'antwerp', 'athens', 'bagus', 'beige', 'bibir', 'bisa', 'blue', 'bold', 'bright', 'brownish', 'cantik', 'cerah', 'coklat', 'color', 'colour', 'coral', 'cukup', 'dan', 'dark', 'deep', 'di', 'dusty', 'elegant', 'finish', 'fuschia', 'gitu', 'gold', 'hasilnya', 'hint', 'hitam', 'ini', 'jadi', 'juga', 'keluar', 'lama', 'lebih', 'lembut', 'light', 'lip', 'lipstick', 'lipsticknya', 'lipstik', 'lumayan', 'malah', 'matte', 'mauve', 'merah', 'merahnya', 'muda', 'muted', 'natural', 'ngejreng', 'nude', 'nya', 'opaque', 'orange', 'orangnya', 'pas', 'peach', 'peachy', 'perfect', 'pigmented', 'pink', 'pinkish', 'pinknya', 'plum', 'pucat', 'red', 'rose', 'ruby', 'russian', 'sedikit', 'sekali', 'semimatte', 'shade', 'soft', 'stockholm', 'tapi', 'terlalu', 'tua', 'ungu', 'velvet', 'warm', 'warna', 'warnanya', 'yang'	33,08%
I-ASP-MOI	'agak', 'bibir', 'buttery', 'creamy', 'creme', 'cukup', 'dan', 'dry', 'extra', 'finish', 'finishing', 'formula', 'glossy', 'hasilnya', 'ini', 'jadi', 'juga', 'kering', 'lembab', 'lembut', 'light', 'lip', 'lips', 'lipstick', 'lipsticknya', 'liquid', 'longlasting', 'malah', 'matte', 'mattenya', 'melembabkan', 'moist', 'moisturizing', 'nya', 'pecah', 'pigmented', 'satin', 'sedikit', 'sekali', 'semi', 'shade', 'smooth', 'super', 'tapi', 'tekstur', 'tetep', 'texture', 'tidak', 'yang'	68,57%
I-ASP-PGM	'cukup', 'dan', 'keluar', 'kurang', 'lips', 'opaque', 'pigmentasi', 'pigmented', 'sangat', 'sekali', 'sendiri', 'super', 'warnanya'	76,47%
I-ASP-PKG	'aplikatornya', 'bagus', 'bentuknya', 'berwarna', 'cantik', 'color', 'dan', 'di', 'elegant', 'gampang', 'gold', 'hitam', 'ini', 'juga', 'kemasannya', 'kurang', 'lasting', 'lipstick', 'lipsticknya', 'lipstik', 'matte', 'merah', 'mudah', 'nya', 'packaging', 'packagingnya', 'pas', 'pecah', 'produk', 'rose', 'semua', 'sendiri', 'terlalu', 'terlihat', 'warna', 'yang'	60,34%
I-ASP-PRI	'affordable', 'gitu', 'harga', 'harganya', 'ini', 'lebih', 'lumayan', 'mahal', 'murah', 'nya', 'price', 'sama', 'sekali', 'tidak', 'yang'	65,22%
I-ASP-STA	'awet', 'awetnya', 'bibir', 'bisa', 'di', 'gampang', 'hilang', 'juga', 'ketahanan', 'lama', 'lasting', 'long', 'mudah', 'nempel', 'nya', 'pudar', 'sekali', 'stay', 'staying', 'tahan', 'tidak', 'walaupun', 'warnanya', 'yang'	50%

Dalam tabel 17 terdapat kolom '*Error rate*' yang menampilkan persentase kemungkinan kesalahan pelabelan oleh sistem yang dibangun. *Error Rate* didapatkan dari persentase jumlah variasi token multi label dibandingkan dengan total variasi token dari label tersebut yang didapatkan pada tabel 6. Berdasarkan tabel 18 diatas, label I-ASP-PGM, I-ASP-MOI, I-ASP-PRI memiliki persentase *error rate* yang tinggi.

## 5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang dapat diambil dari penelitian ini sebagai berikut. Arsitektur sistem yang dibangun atas BiLSTM-CRFs ini mendapatkan hasil terbaik dengan rata-rata *F1 Score* sebesar 57% pada uji skenario 4 dengan parameter dimensi vektor *embedding layer* sebesar 256, banyak *batch* sebesar 32, dan banyak *epoch* sebesar 64. Hasil tersebut melebihi kinerja dari model LSTM-CRFs terbaik yang dilakukan pada penelitian ini, dengan perolehan rata-rata *F1 Score* sebesar 56%. Banyaknya label pada setiap aspek, dan variasi token aspek berpengaruh besar pada pelatihan sistem. Semakin bervariasi token aspek maka akan semakin sulit dalam pengklasifikasian token tersebut pada data uji.

Nilai *F1 Score* untuk label B-ASP-MOI, I-ASP-PKG, I-ASP-PRI cenderung paling rendah dalam setiap pengujian. Hal ini dapat dijelaskan dengan analisis error yang telah dilakukan, dimana ketiga label tersebut mempunyai banyak variasi token dengan label yang hanya terdapat pada data uji, dan token-token dalam label tersebut mempunyai label lebih dari satu label.

Saran untuk penelitian selanjutnya yaitu memperbanyak dataset yang dapat berpengaruh pada persebaran label pada data uji dan data latih sehingga dapat meningkatkan hasil prediksi dari sistem yang dibuat. Arsitektur *layer* dapat dimodifikasi sehingga lebih efisien dan mengurangi *running time program*.



**REFERENSI**

- [1] Pakdosen, (2020). Pengertian Teks Ulasan. Tersedia pada: <https://pakdosen.co.id/teks-ulasan> (diakses Februari 14, 2020).
- [2] Brightlocal, (2021). Local Consumer Review Survey 2020. Rosie Murphie. Tersedia pada: <https://www.brightlocal.com/research/local-consumer-review-survey> (diakses Juli 01, 2021)
- [3] Fernando, Jordhy & Khodra, Masayu & Septiandri, Ali. (2019). Aspect and Opinion Terms Extraction Using Double Embeddings and Attention Mechanism for Indonesian Hotel Reviews.
- [4] Poria, S., Cambria, E., & Gelbukh, A. (2019) .Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*. 108, 42–49. doi: 10.1016/j.knsys.2016.06.009
- [5] I Gede Agus Maha Putra. (2017). Pengembangan Ekstraksi Fitur dan Opini Produk Menggunakan Pendekatan Pattern Knowledge dan Taxonomy. Telkom University, 17.04.3574
- [6] Lê, Thê Anh & Arkhipov, Mikhail & Burtsev, Mikhail. (2018). Application of a Hybrid Bi-LSTM-CRF Model to the Task of Russian Named Entity Recognition. 91-103. 10.1007/978-3-319-71746-3\_8.
- [7] Giannakopoulos, Athanasios & Musat, Claudiu & Hossmann, Andreea & Baeriswyl, Michael. (2017). Unsupervised Aspect Term Extraction with B-LSTM and CRF using Automatically Labelled Datasets. 10.18653/v1/W17-5224.
- [8] Tao Chen, Ruifeng Xu, Yulan He, Xuan Wang. (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications*, Volume 72, 2017, Pages 221-230, ISSN 0957-4174.
- [9] M. Tubishat, N. Idris, and M. A. Abushariah. (2018). Implicit aspect extraction in sentiment analysis. *Information Processing and Management: an International Journal* Volume 54 Issue 4 July 2018 pp 545–563
- [10] Miao, Ya Lin et al. (2021). Aspect-based Sentiment Analysis in Chinese Based on Mobile Reviews for BiLSTM-CRF. *Journal of Intelligent & Fuzzy Systems (IF)*, Pub Date : 2021-02-16, DOI: 10.3233/jifs-192078
- [11] Poria, Soujanya & Cambria, Erik & Ku, Lun-Wei & Gui, Chen & Gelbukh, Alexander. (2014). A Rule-Based Approach to Aspect Extraction from Product Reviews. *SocialNLP*. 2014. 10.3115/v1/W14-5905.
- [12] Mukherjee, Arjun & Liu, Bing. (2012). Aspect extraction through semi-supervised modeling. 50th Annual Meeting of the Association for Computational Linguistics. *ACL 2012 - Proceedings of the Conference*. 1. 339-348.
- [13] Ashish Bansal. (2021). *2021 Advanced Natural Language Processing Using TensorFlow 2*. Birmingham – Mumbai, Packt. ISBN 978-1-80020-093-7.

