Analisis Sentimen Level Topik Berbasis *Large Language Model* GPT pada Ulasan Aplikasi *E-groceries* di Indonesia

Safira Raissa Rahmi
2006568891

Dosen Pembimbing:
Dr. rer. nat. Hendri Murfi, S.Si., M.Kom.
Dra. Nora Hariadi, M.Si.

Dosen Penguji:
Drs. Gatot Fatwanto Hertono, M.Sc., Ph.D.
Gianinna Ardaneswari, M.Si.

Program Studi S1 Matematika Departemen Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) - Universitas Indonesia Gedung D, Kampus UI, Depok 16424,
Telp. +62-21-7862719, Fax. +62-21-7863439, Email. sekretariat.math@sci.ui.ac.id

Agenda Presentasi

1. Pendahuluan

2. Metode Analisis Sentimen pada Level Topik

3. Simulasi dan Analisis Hasil

4. Penutup

Pendahuluan

Latar Belakang
Rumusan Masalah
Tujuan Penelitian
Batasan Masalah

Latar Belakang: *E-Groceries*

Total *Gross Merchandise Value* (**GMV**) Indonesia menunjukkan **tren positif** sejak pandemi COVID-19 pada 2020.

Kenaikan GMV mencapai **US$ 82 miliar** pada **2023, 75,61%** sektor *e-commerce*

High Value Users (**HVU**) *e-commerce* pada kategori bahan makanan (***groceries***) meningkat **7,2 kali** lipat dibandingkan tahun 2022.

Diperlukan **analisis ulasan pengguna *e-groceries*** di Indonesia.

(Sumber data: e-Conomy SEA)

E-groceries merupakan **sektor *e-commerce*** yang berfokus pada penjualan **bahan makanan** secara ***online***, menawarkan kenyamanan dan pengiriman cepat (Jagani, K. *et al*., 2020).

Latar Belakang: Analisis Sentimen

Analisis ulasan pengguna ***e-groceries*** di Indonesia.

Analisis Sentimen

Metode:

Konvensional

- CNN, LSTM, dan GRU (Zhang, *et al.*, 2018)

Large Languange Model

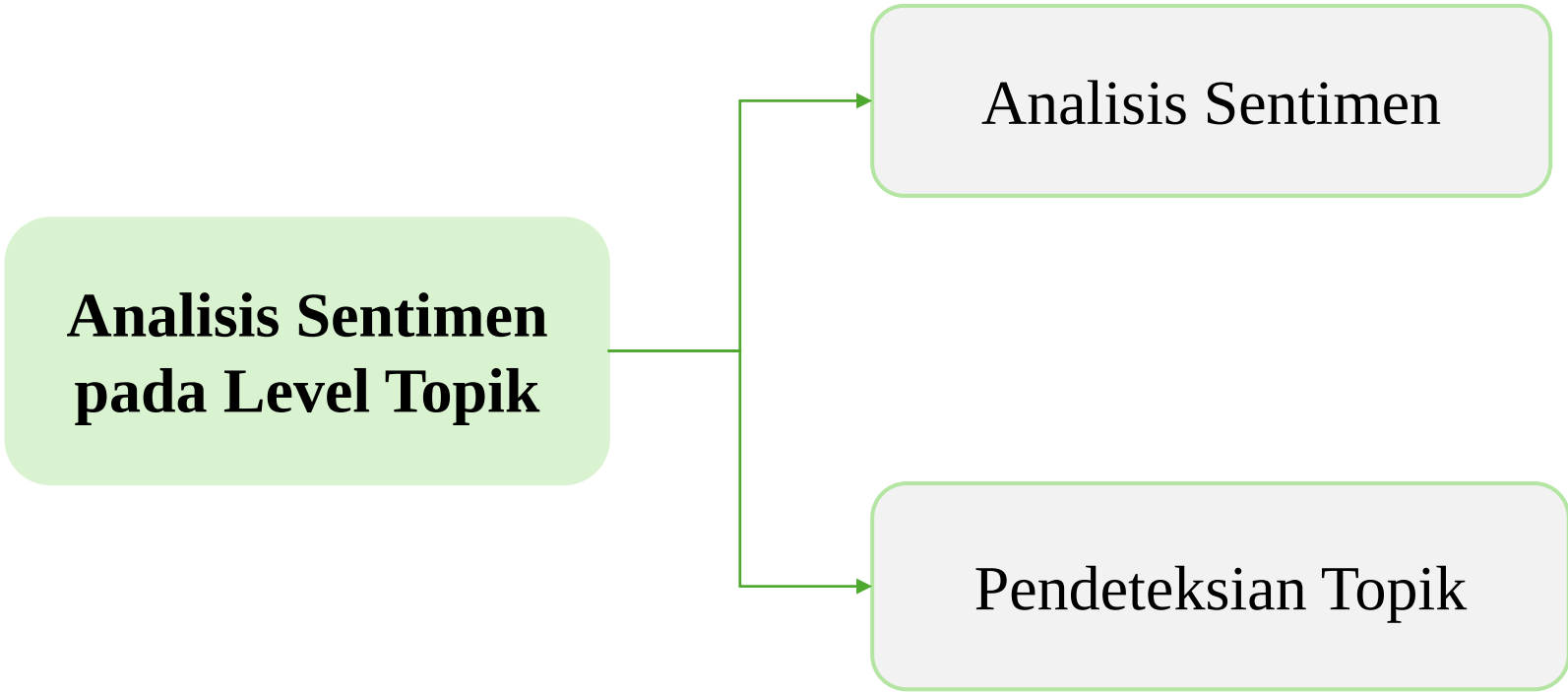
- RoBERTa (Nguyen, *et al.*, 2020)
- *Generative Pretrained Model* (**GPT**) (Kheiri & Karimi, 2023).

Umumnya hanya pada **level dokumen**

Analisis Sentimen pada Level Topik, sebagai pendekatan yang lebih terperinci

Analisis sentimen merupakan bidang studi yang mengeksplorasi **opini**, **emosi**, dan **penilaian** terhadap **suatu entitas** (Liu, 2012)

Latar Belakang: Analisis Sentimen pada Level Topik



Latar Belakang: Pendeteksian Topik

Analisis Sentimen pada Level Topik

→

Pendeteksian Topik

→

Metode:
Representasi Teks
+
Clustering

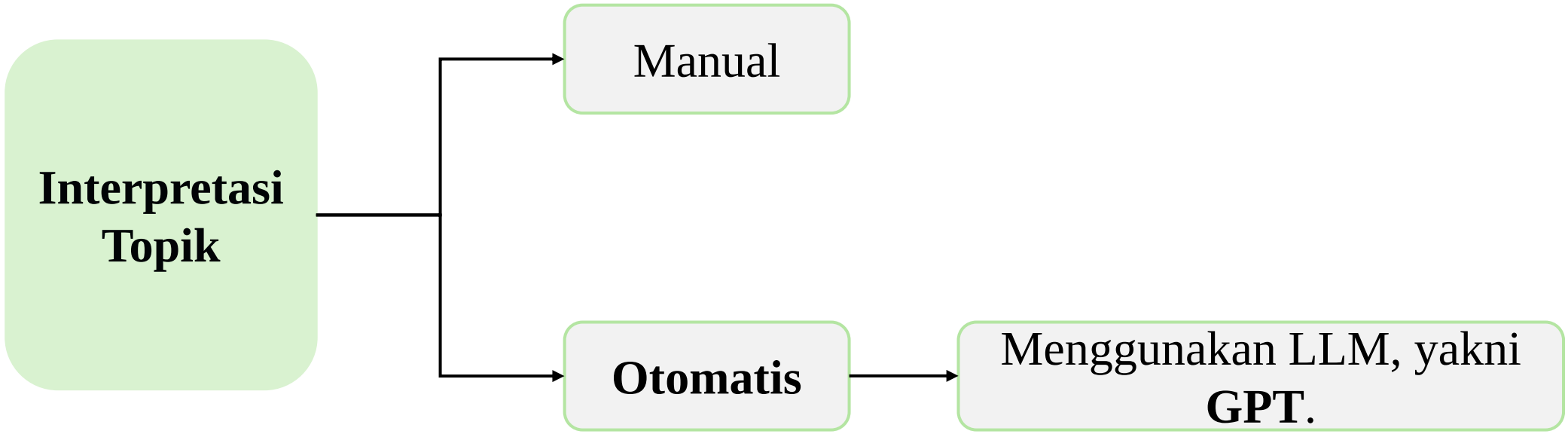
→

BERT-EFCM:

- BERT merupakan LLM yg telah mencapai *state of art.*
- EFCM memiliki skor koherensi antara LDA dan NMF (Murfi, 2021).
- Performa BERT-EFCM dengan lebih unggul dibandingkan TFIDF-EFCM. (Yudhistira *et al*., 2023)

Pendeteksian topik merupakan metode untuk menentukan topik secara otomatis (Garcia & Berton, 2021).

Latar Belakang: Pendeteksian Topik



Rumusan Masalah

1. Apa saja topik-topik utama pada ulasan pengguna aplikasi *e-groceries* di Indonesia?

2. Bagaimana sentimen pengguna aplikasi *e-groceries* di Indonesia terhadap topik-topik tersebut?

Tujuan Penelitian

1. Mendeteksi topik-topik utama pada ulasan aplikasi *e-groceries* di Indonesia menggunakan model BERT-EFCM dan merepresentasikan topik-topik utama tersebut melalui automasi interpretasi topik menggunakan model GPT.

2. Mengklasifikasi sentimen untuk setiap topik-topik utama tersebut menggunakan model GPT.

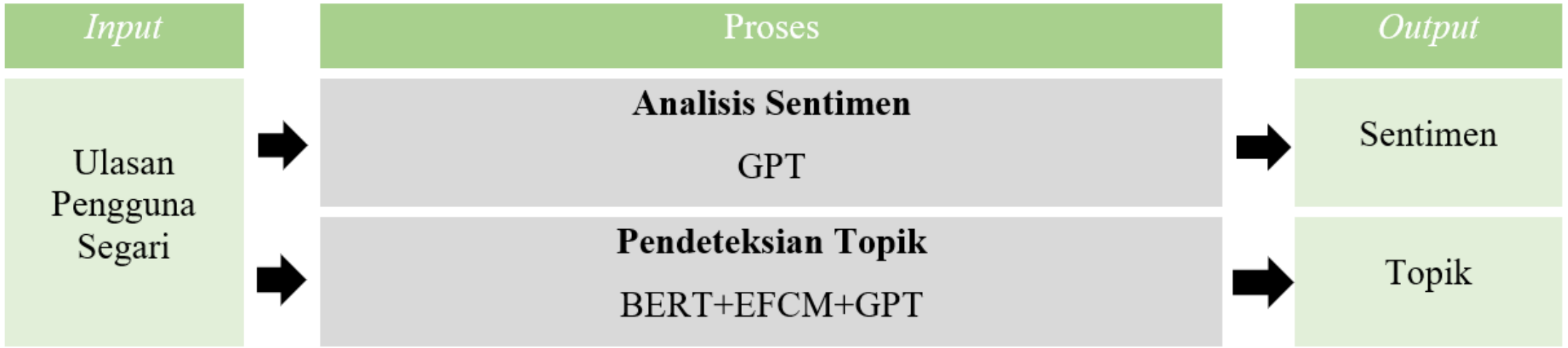
Batasan Masalah

1.	Data yang digunakan adalah data ulasan konsumen terhadap aplikasi *e-groceries* di Indonesia dengan jumlah ulasan terbanyak pada aplikasi Play Store, yakni Segari.
2.	Data yang diambil adalah data dari rentang waktu 31 Desember 2022 sampai 31 Desember 2023.
3.	Ulasan yang digunakan hanya ulasan dengan bahasa Indonesia.
4.	Ulasan yang diambil merupakan 3.078 ulasan paling relevan dari aplikasi *e-groceries* yang telah disebutkan berdasarkan Play Store.
5.	Sentimen yang dianalisis berupa sentimen positif dan negatif.
6.	Model GPT yang digunakan dalam penelitian ini adalah model GPT-3.5 Turbo.

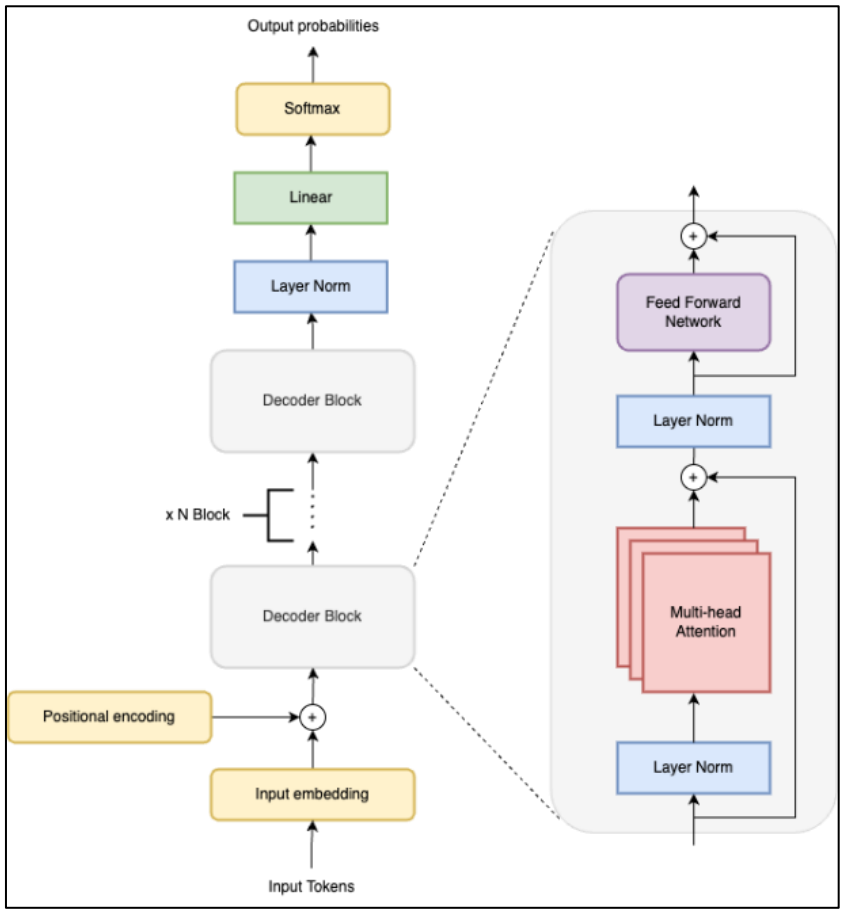
Metode Analisis Sentimen pada Level Topik

Metode Analisis Sentimen
Metode Pendeteksian Topik

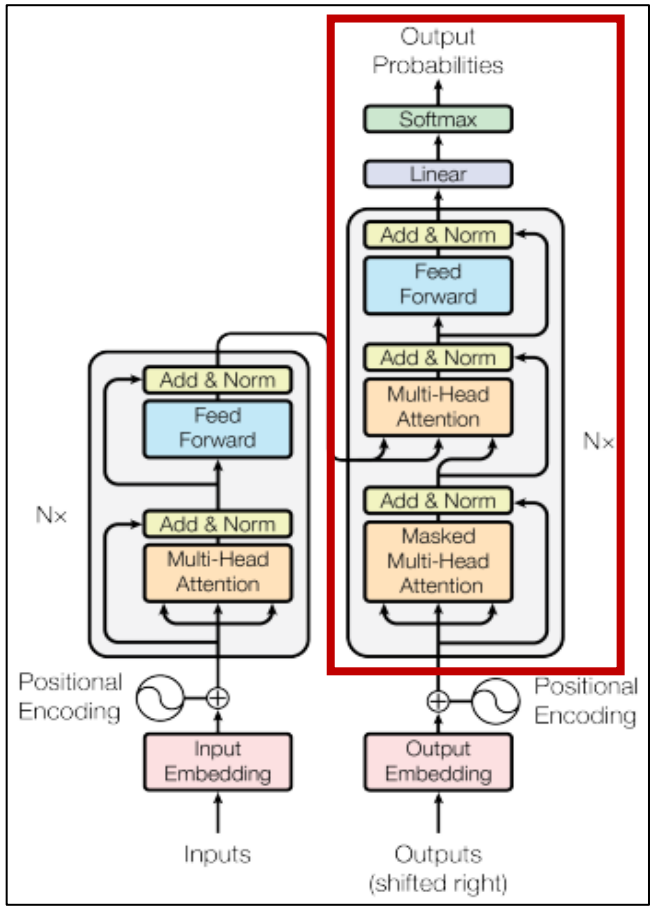
Alur Proses Keseluruhan



Metode Analisis Sentimen: *Generative Pretrained Transformer* (GPT)



Arsitektur model *decoder-only*
(Le *et. al*., 2023)



Arsitektur *Transformer*
(Vaswani et al., 2017)

GPT merupakan model bahasa yang dikembangkan oleh OpenAI dengan menggunakan arsitektur *decoder-only transformer* (Radford *et al*., 2018).

GPT dapat menghasilkan teks mirip bahasa manusia secara **autoregresif** berdasarkan *input* teks yang diberikan.

Metode Analisis Sentimen: Representasi *Input* GPT

Tokenisasi

Membagi kalimat menjadi bagian-bagian lebih kecil yang disebut **token**. Dilakukan menggunakan model ***Byte Pairs Encoding* (BPE)**.

Kalimat Awal	"aplikasi *e-groceries* terlengkap, pengirimannya cepat"
Tokenisasi	["ap", "lik", "asi", "e", "-g", "ro", "ceries", "ter", "l", "engkap", ",", "peng", "irim", "annya", "cep", "at"]
Penambahan *Special Token*	[<SOS>, "ap", "lik", "asi", "e", "-g", "ro", "ceries", "ter", "l", "engkap", ",", "peng", "irim", "annya", "cep", "at", <EOS>]

Metode Analisis Sentimen: Representasi *Input* GPT

Embedding

Memetakan token menjadi representasi **vektor numerik**. Tahapannya terdiri dari dua langkah.

Input	<SOS>	"ap"	"lik"	"asi"	"e"	"-g"	"ro"	"ceries"	"ter"	"l"	"engkap"	","	"peng"	"irim"	"annya"	"cep"	"at"	<EOS>
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Embedding	$E_{<SOS>}$	E_{ap}	E_{lik}	E_{asi}	E_{e}	$E_{\_g}$	E_{ro}	E_{ceries}	E_{ter}	E_{l}	E_{engkap}	$E_{,}$	E_{peng}	E_{irim}	E_{annya}	E_{cep}	E_{at}	$E_{<EOS>}$
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Positional Embedding	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}	E_{13}	E_{14}	E_{15}	E_{16}	E_{17}

(Radford *et al.*, 2018; telah diolah kembali)

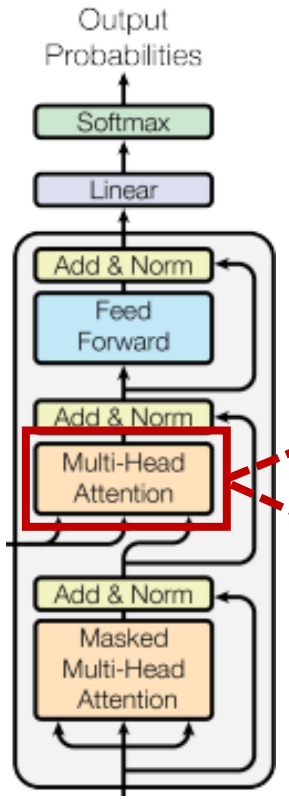
Metode Analisis Sentimen: Representasi *Input* GPT

Embedding

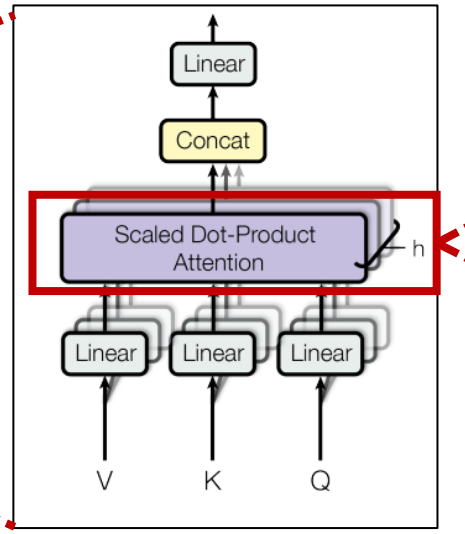
token	***token embedding***		***position embedding***		**representasi *input***
<SOS>	(0.15, …, -0.30)	+	(-0.11, …, -0.11)	=	(0.04, …, -0.41)
“ap”	(0.28, …, 0.33)	+	(-0.01, …, -0.88)	=	(0.27, …, -0.55)
“lik”	(-0.03, …, 0.22)	+	(-0.98, …, -0.20)	=	(-1.11, …, 0.02)
“asi”	(-0.05, …, -0.45)	+	(-0.26, …, 0.27)	=	(-0.31, …, -0.18)
	ukuran=12.288		ukuran=12.288		ukuran=12.288

Metode Analisis Sentimen: *Decoder*

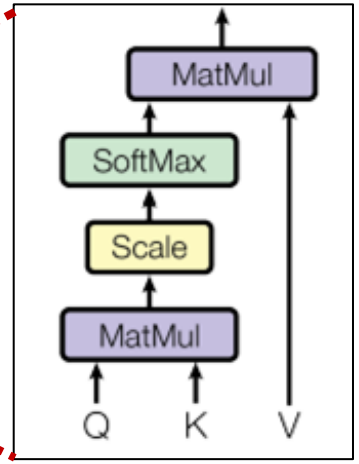
Decoder : *Attention*



Arsitektur *Decoder*
(Vaswani *et al.*, 2017)



Ilustrasi *Multi-Head Attention*
(Vaswani *et al.*, 2017)

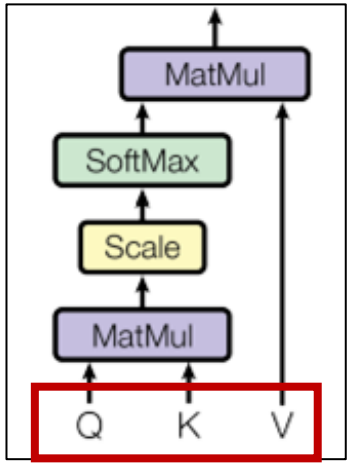


Ilustrasi *Scaled Dot-Product Attention*
(Vaswani *et al.*, 2017)

Attention merupakan fungsi pemetaan nilai perhatian suatu kata ke kata lain (Vaswani, 2017).

Metode Analisis Sentimen: *Decoder*

Decoder: Scaled Dot-Product Attention

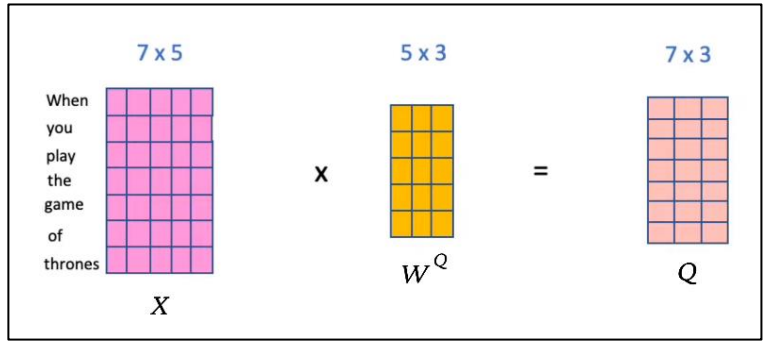


Ilustrasi *Scaled Dot-Product Attention*
(Vaswani *et al.*, 2017)

Scaled dot-product attention menerima tiga *input* yakni Q matriks ***query*** berdimensi d_q, K matriks ***key*** berdimensi $d_{k,}$, dan V matriks ***value*** berdimensi d_v.

$$Q = XW^Q$$
$$K = XW^K$$
$$V = XW^V$$

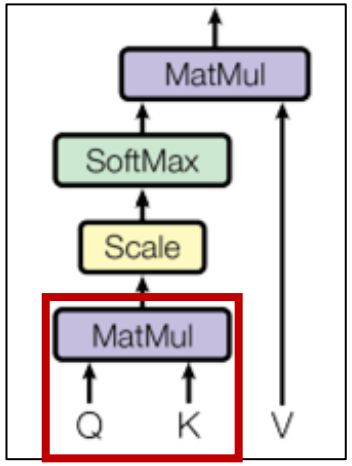
Keterangan:
Matriks bobot $W^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W^K \in \mathbb{R}^{d_{model} \times d_k}$, dan $W^V \in \mathbb{R}^{d_{model} \times d_v}$ merupakan hasil dari proses pelatihan model



Ilustrasi pencarian matriks *Q*
(Haider, 2020), telah diolah kembali

Metode Analisis Sentimen: *Decoder*

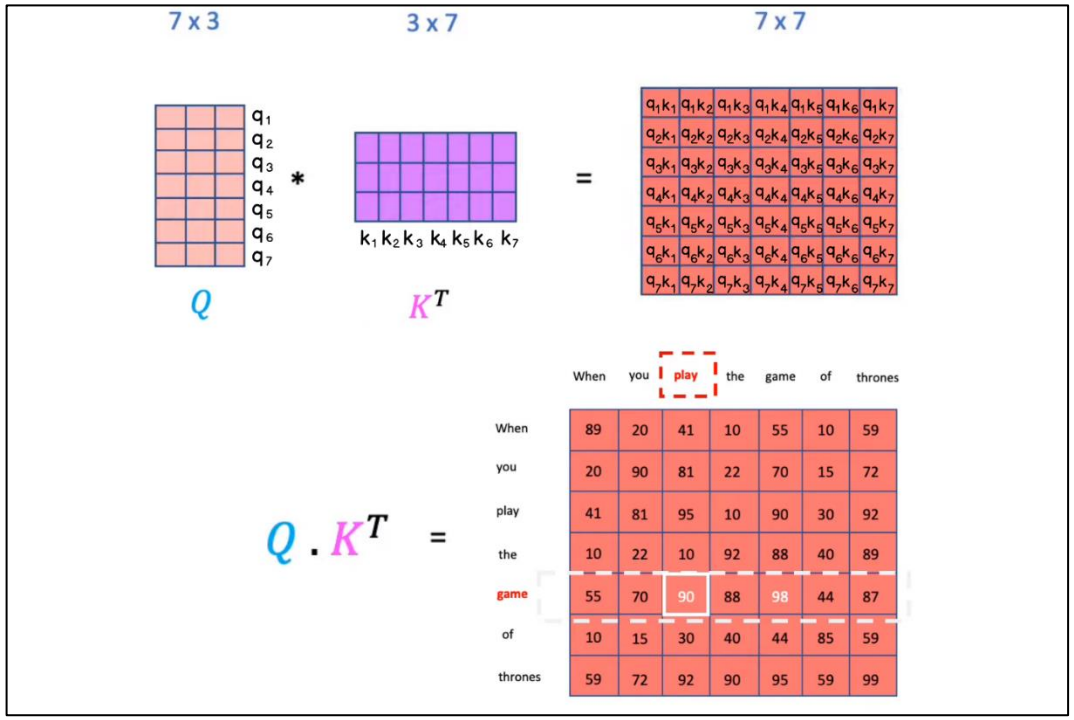
Decoder: Scaled Dot-Product Attention



Ilustrasi *Scaled Dot-Product Attention*
(Vaswani *et al.*, 2017)

1.Menghitung *matrix multiplication* antara matriks Q dengan matriks K^T

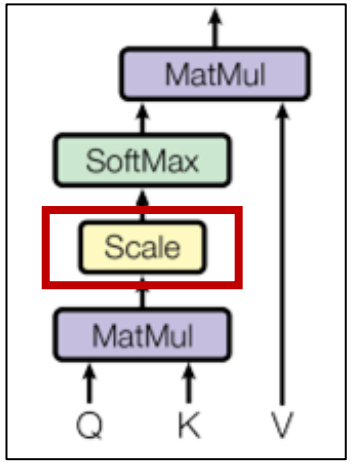
Tujuan : Mengecek keterkaitan satu kata dengan kata lainnya pada suatu kalimat atau dokumen



Ilustrasi perkalian matriks Q dan K^T
(Haider, 2020), telah diolah kembali

Metode Analisis Sentimen: *Decoder*

Decoder: Scaled Dot-Product Attention



Ilustrasi *Scaled Dot-Product Attention*
(Vaswani *et al.*, 2017)

2. Melakukan *Scaling*

Melakukan **pembagian** nilai $\boldsymbol{QK^T}$ dengan akar dimensi vektor *key* yaitu $\sqrt{\boldsymbol{d_k}}$

Tujuan : Mendapatkan gradien model yang stabil.

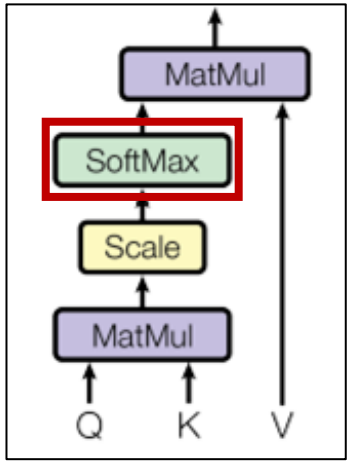
$$\frac{Q \cdot K^T}{\sqrt{d_k}} = \frac{Q \cdot K^T}{\sqrt{7}} =$$

	When	you	play	the	game	of	thrones
When	33.6	7.6	15.5	3.8	20.8	3.8	22.3
you	7.6	34.0	30.6	8.3	26.5	5.7	27.2
play	15.5	30.6	35.9	3.8	34.0	11.3	34.8
the	3.8	8.3	3.8	34.8	33.3	15.1	33.6
game	20.8	26.5	34.0	33.3	37.0	16.6	35.9
of	3.8	5.7	11.3	15.1	16.6	32.1	22.3
thrones	22.3	27.2	34.8	34.0	35.9	22.3	37.4

Ilustrasi perhitungan *scaling*
(Haider, 2020), telah diolah kembali

Metode Analisis Sentimen: *Decoder*

Decoder: Scaled Dot-Product Attention



Ilustrasi *Scaled Dot-Product Attention*
(Vaswani *et al.*, 2017)

3. Melakukan Normalisasi dengan Fungsi *Softmax*

Tujuan : Membantu model mengetahui keterhubungan suatu kata dengan seluruh kata pada kalimat.

$$\text{Softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) =$$

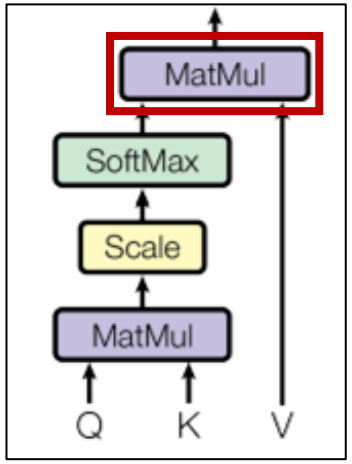
	When	you	play	the	game	of	thrones
When	1.00	0.00	0.00	0.00	0.00	0.00	0.00
you	0.00	0.97	0.03	0.00	0.00	0.00	0.00
play	0.00	0.00	0.68	0.00	0.10	0.00	0.22
the	0.00	0.00	0.00	0.65	0.14	0.00	0.21
game	0.00	0.00	0.03	0.02	0.72	0.00	0.23
of	0.00	0.00	0.00	0.00	0.00	1.00	0.00
thrones	0.00	0.00	0.05	0.03	0.17	0.00	0.75

Normalisasi dengan fungsi *softmax*
(Haider, 2020), telah diolah kembali

Metode Analisis Sentimen: *Decoder*

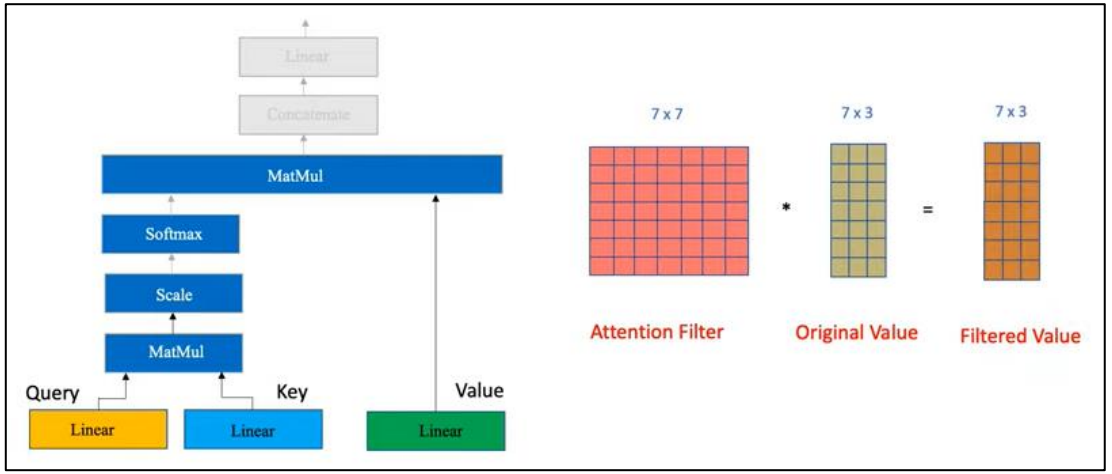
Decoder: Scaled Dot-Product Attention

4. Menghitung Nilai *Attention*



Ilustrasi *Scaled Dot-Product Attention*
(Vaswani *et al.*, 2017)

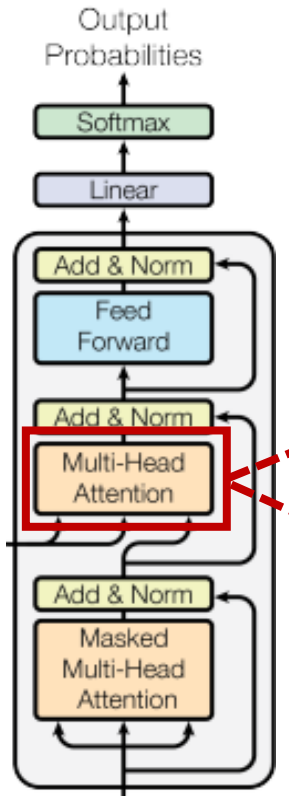
$$Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



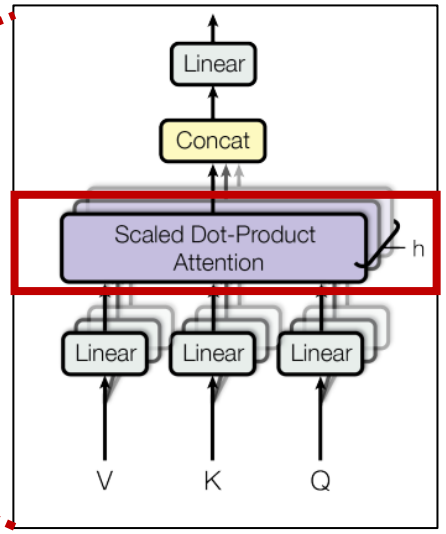
Ilustrasi dan intuisi matriks *attention*
(Haider, 2020), telah diolah kembali

Metode Analisis Sentimen: *Decoder*

Decoder : *Multi-Head Attention*



Arsitektur *Decoder*
(Vaswani *et al.*, 2017)



Ilustrasi *Multi-Head Attention*
(Vaswani *et al.*, 2017)

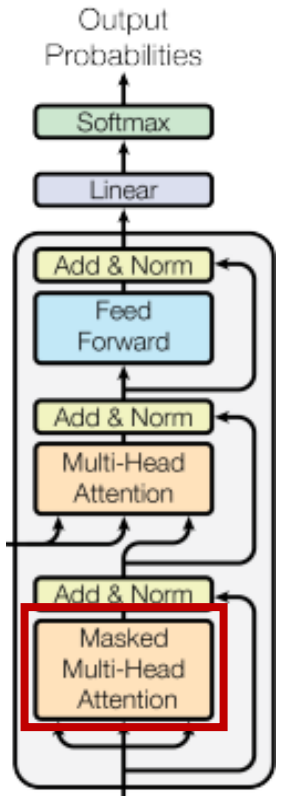
$$Z_j = Attention(Q_j, K_j, V_j) = Softmax\left(\frac{Q_j {K_j}^T}{\sqrt{d_k}}\right) V_j$$

$$Multihead\ attention(Q, K, V) = Concatenate(Z_1, \dots, Z_h) W^O$$

Keterangan:
$W^O \in \mathbb{R}^{hd_v \times d_{model}}$ merupakan matriks bobot
Z_j matriks *attention* dari *head* ke- j

Multi-head attention merupakan **penggabungan *self-attention* sebanyak *h*** *attention* berbeda yang berjalan secara simultan

Metode Analisis Sentimen: *Decoder*

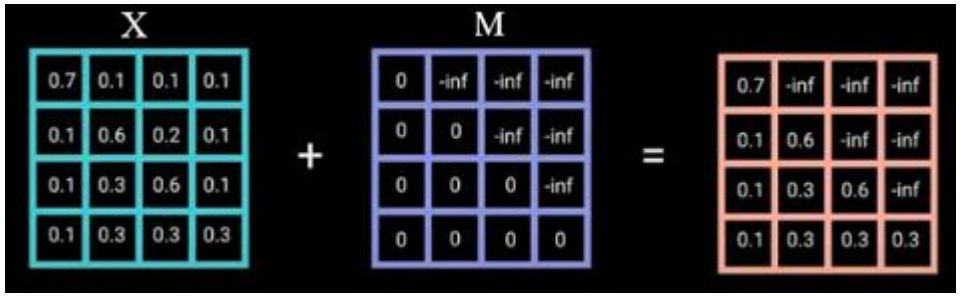


Arsitektur *Decoder*
(Vaswani *et al.*, 2017)

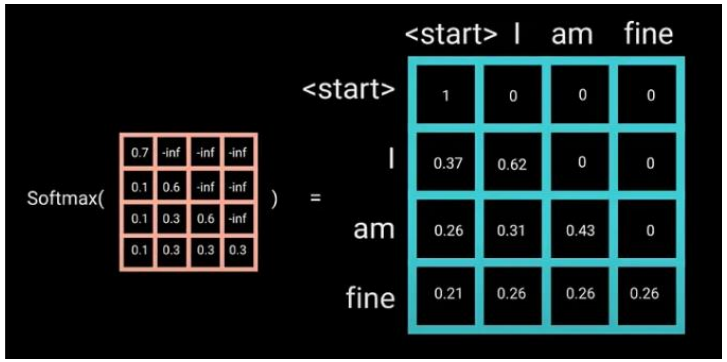
Decoder : *Masked Multi-Head Attention*

Tujuan: Mencegah *decoder* melihat kata-kata selanjutnya dalam sekuens *output*.

$$Masked\ Attention(X, M, V) = Softmax\left(\frac{X + M}{\sqrt{d_k}}\right)V$$

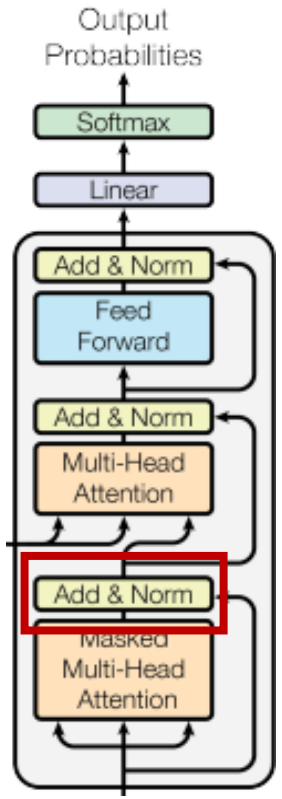


Ilustrasi proses *masked attention* sebelum *softmax* (Phi, 2020; telah diolah kembali)



Ilustrasi penerapan *softmax* pada matriks *mask attention* pada *decoder*
(Phi, 2020; telah diolah kembali)

Metode Analisis Sentimen: *Decoder*



Arsitektur *Decoder*
(Vaswani *et al.*, 2017)

Decoder : Koneksi Residu dan Normalisasi

Koneksi Residu

Tujuan: Mengurangi risiko kehilangan informasi dari data *input* pada saat *training* seiring bertambahnya kedalaman arsitektur model (He *et al.*, 2016).

$$Z' = X + Z$$

Keterangan:
- X *input* dan out*put* Z *lapisan sebelumnya*
- g adalah parameter *gain*
- μ dan σ adalah rata-rata dan standar deviasi dari H neuron dari *hidden layer* h_i

Normalisasi

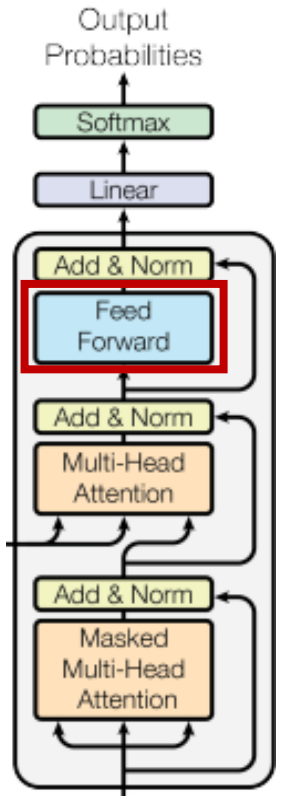
Tujuan : Mengatasi *covariate-shift*, menormalisasi *output* neuron menjadi distribusi normal (μ=0 & σ=1).

$$h_i = \frac{g}{\sigma}(h_i - \mu)$$

$$\mu = \frac{1}{H}\sum_{i=1}^{H} h_i$$

$$\sigma = \sqrt{\sum_{i=1}^{H}(h_i - \mu)^2}$$

Metode Analisis Sentimen: *Decoder*



Arsitektur *Decoder*
(Vaswani *et al.*, 2017)

Decoder : *Position-wise Feed Forward Network*

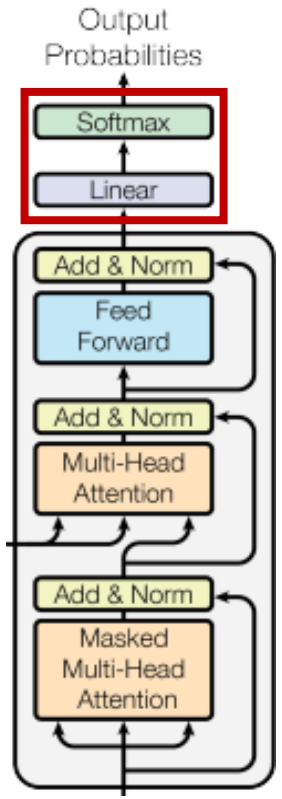
Tujuan : Melakukan transformasi kata pada setiap posisi dalam dokumen menggunakan FFN yang sama (Zhang *et al.*, 2017).

Terdiri dari dua transformasi linier dan diantaranya digunakan fungsi aktivasi ReLU.

$$FFN(X) = max(0, XW_1 + b_1)\ W_2 + b_2$$

Keterangan:
- X adalah vektor *input*
- W_1 dan W_2 secara berurutan adalah matriks parameter bobot untuk lapisan ke-1 (lapisan tersembunyi) dan ke-2 (lapisan *output*)
- b_1 dan b_2 adalah parameter bias

Metode Analisis Sentimen: *Decoder*



Arsitektur *Decoder*
(Vaswani *et al.*, 2017)

Decoder : *Linear* dan *Softmax Layer*

Linear

Tujuan : Mengubah dimensi *hidden state* menjadi sesuai dengan ukuran kosa kata keluaran (*vocabulary size*). Hal ini agar data dapat diolah lebih lanjut oleh *softmax layer*.

$$Z = WY + b$$

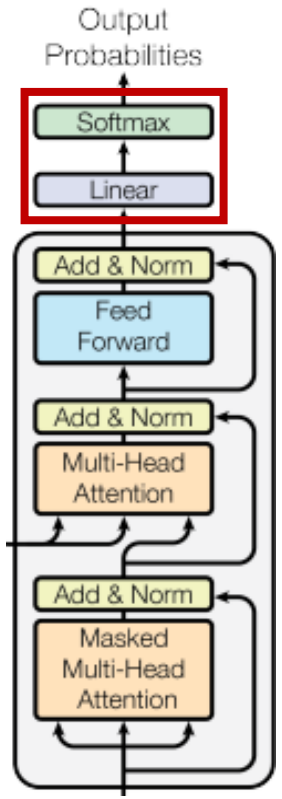
Softmax

Tujuan : Mengkonversi *output linear layer* menjadi distribusi probabilitas atas kosa kata keluaran pada *softmax layer.*

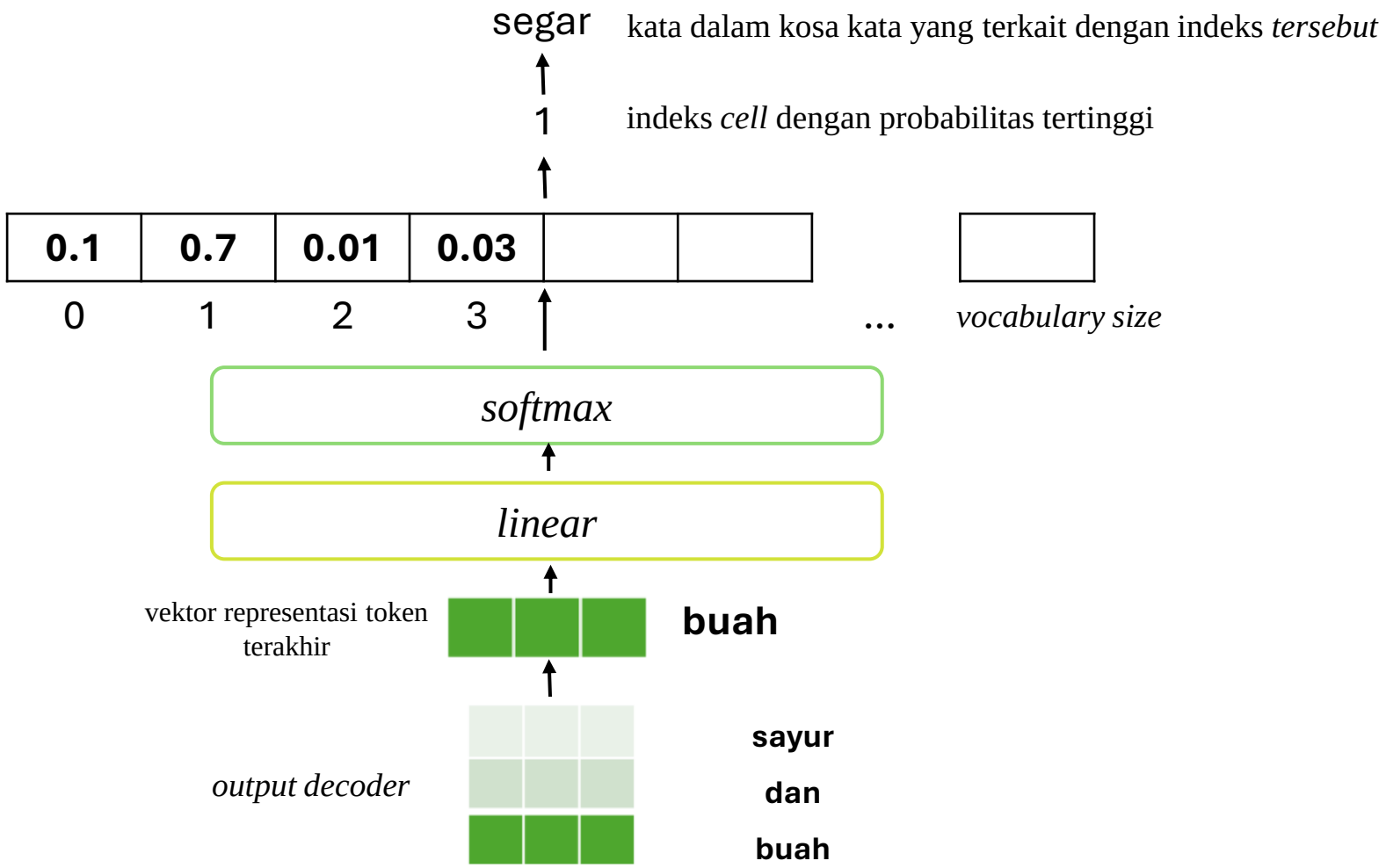
$$softmax(z_i) = \frac{exp(z_i)}{\sum_{j=1}^{v} exp(z_j)}$$

Keterangan:
- z_i adalah elemen ke-i dari vektor Z
- v adalah ukuran *vocabulary size*, $j = 1, \ldots, v$ adalah indeks yang mewakili setiap token dalam kosa kata

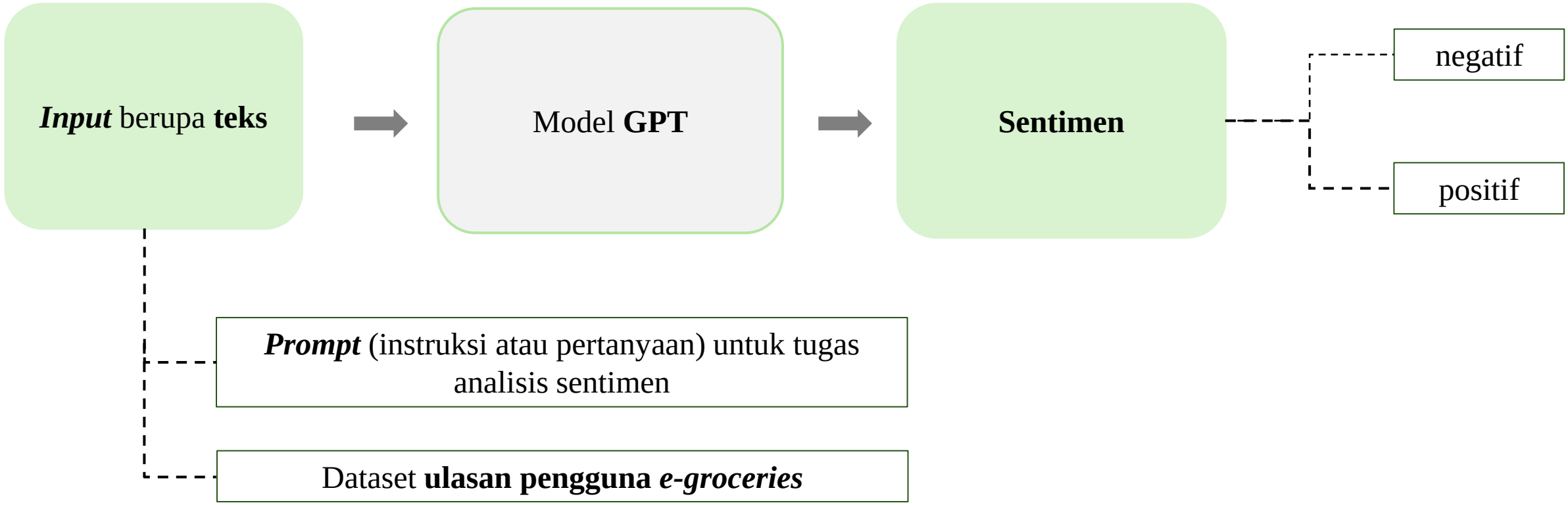
Metode Analisis Sentimen: *Decoder*



Arsitektur *Decoder*
(Vaswani *et al.*, 2017)



Metode Analisis Sentimen: GPT untuk Analisis Sentimen



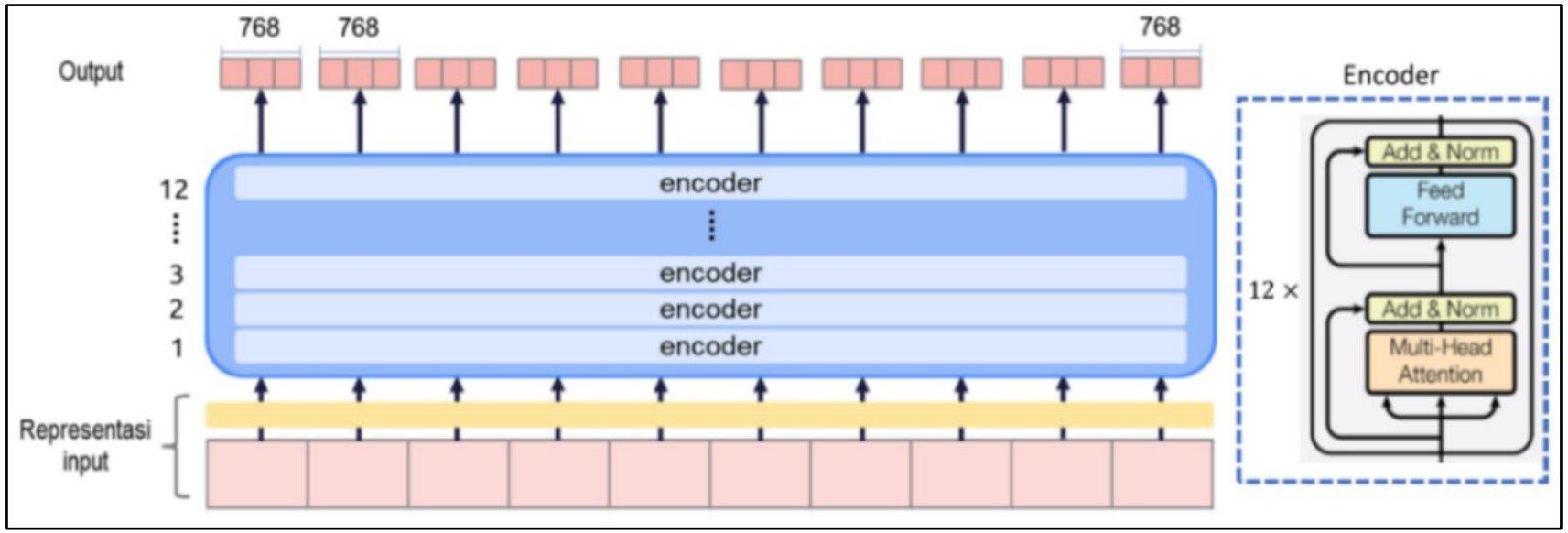
Metode Analisis Sentimen: GPT untuk Analisis Sentimen

Contoh *Prompt* untuk Tugas Analisis Sentimen

Saya sedang melakukan analisis sentimen terhadap ulasan pelanggan dalam bidang *e-groceries*. Tolong analisis ulasan berikut dan tentukan apakah sentimennya positif atau negatif. Petunjuk:
1. *E-groceries* adalah layanan belanja bahan makanan secara *online* yang menawarkan kemudahan dan kecepatan dalam memenuhi kebutuhan sehari-hari pelanggan.
2. Jawab hanya dengan satu kata saja: “positif” atau “negatif”!
3. Jangan menambahkan keterangan apapun pada jawaban. Pastikan hanya satu kata: “positif” atau “negatif”!
<teks>[DATA]</teks>

(Kheiri & Karimi, 2023, telah diolah kembali)

Metode Pendeteksian Topik : *Bidirectional Encoder Representations from Transformer* (BERT)



Arsitektur BERTBASE
(Syamsyuriani, 2021)

BERT adalah model *natural language processing* (NLP) yang memanfaatkan arsitektur *encoder* dari *Transformer* (Devlin *et al.*, 2018).

Metode Pendeteksian Topik : Representasi *Input* BERT

Tokenisasi

Membagi kalimat menjadi bagian-bagian lebih kecil yang disebut **token**. Dilakukan menggunakan model *WordPiece*.

Kalimat Awal	"aplikasi *e-groceries* terlengkap, pengirimannya cepat"
Tokenisasi	[aplikasi, *e-groceries*, terlengkap, pengiriman, ##nya, cepat]
Penambahan *Special Token*	[[CLS], aplikasi, *e-groceries*, terlengkap, [SEP], pengiriman, ##nya, cepat [SEP]]

Metode Pendeteksian Topik : Representasi *Input* BERT

Embedding

Memetakan token menjadi representasi **vektor numerik**. Tahapannya terdiri dari tiga langkah.

Input	[CLS]	aplikasi	e-groceries	terlengkap	[SEP]	pengiriman	##nya	cepat	[SEP]
	↓	↓	↓	↓	↓	↓	↓	↓	↓
Token Embedding	E_{CLS}	$E_{aplikasi}$	$E_{e-groceries}$	$E_{terlengkap}$	E_{SEP}	$E_{pengiriman}$	$E_{\#\#nya}$	E_{cepat}	E_{SEP}
	+	+	+	+	+	+	+	+	+
Segment Embedding	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+
Position Embedding	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8

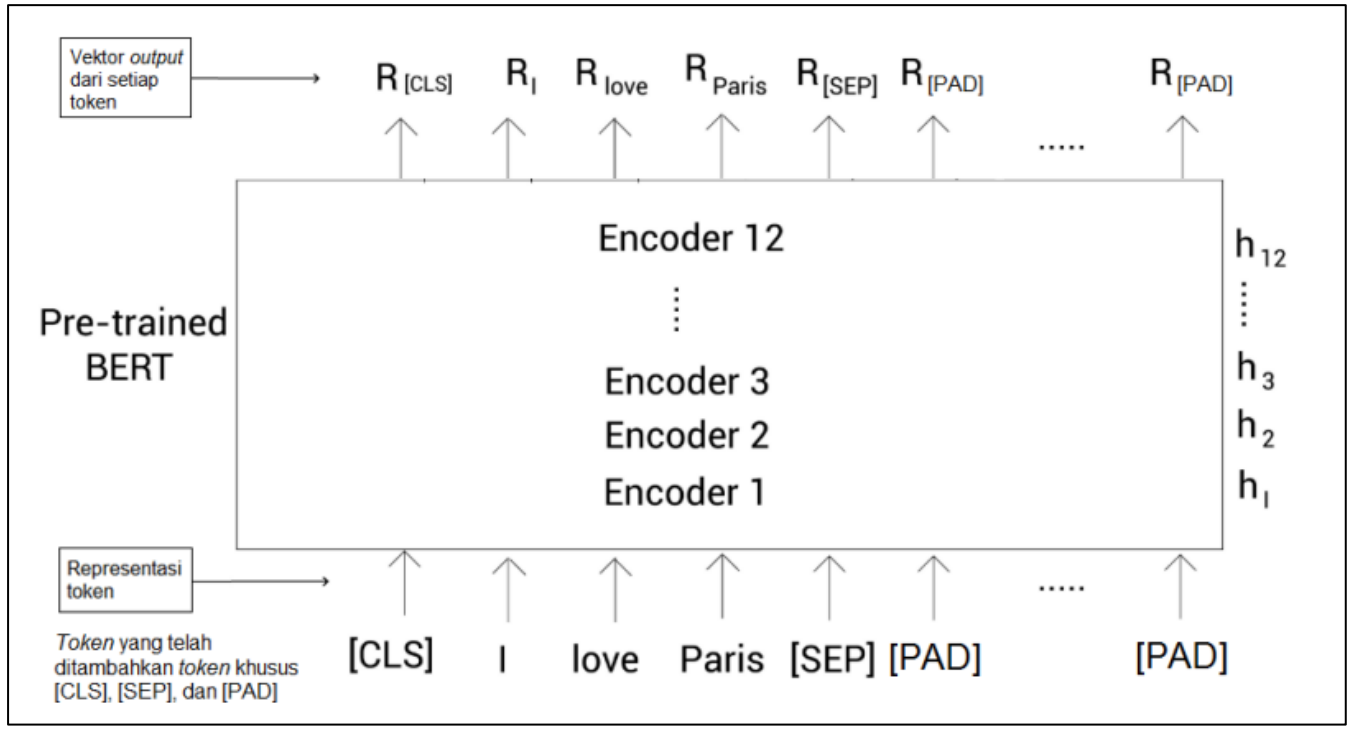
(Devlin *et al*., 2018; telah diolah kembali)

Metode Pendeteksian Topik : Representasi *Input* BERT

Embedding

token	***token embedding***		***segment embedding***		***position embedding***		**representasi *input***
[CLS]	(0.15, …, 0.30)	+	(-0.11, …,-0.30)	+	(0.32, …, -0.01)	=	(0.36, …, -0.01)
aplikasi	(0.28, …, 0.33)	+	(-0.55, …, 0.17)	+	(0.66, …, -0.88)	=	(0.39, …, -0.38)
	ukuran=768		ukuran=768		ukuran=768		ukuran=768

Metode Pendeteksian Topik: BERT sebagai Representasi Teks



Ilustrasi representasi teks menggunakan BERT
(Ravichadiran, 2021), telah diolah kembali

Menggunakan pendekatan ***featured base*** untuk **mengekstrak representasi numerik** dari teks.

Membantu menyelesaikan **berbagai tugas** seperti klasifikasi, *clustering*, dsb.

Digunakan **token khusus [PAD]** untuk menyamakan panjang *input*.

Metode Pendeteksian Topik: EFCM sebagai Metode *Clustering*

Eigenspace-based Fuzzy C-Means (EFCM) merupakan pengoptimalan *Fuzzy C-Means* dengan memanfaatkan ***Truncated* SVD** sebagai **metode reduksi** dimensi pada data.

TSVD

$$\tilde{A}_{m\times n} \approx \tilde{U}_{m\times s}\tilde{\Sigma}_{s\times s}(\tilde{V}_{n\times s})^T \approx [u_1\ u_2 \dots\ u_s]\begin{bmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_s\end{bmatrix}\begin{bmatrix}v_1^T\\ v_2^T\\ \vdots\\ v_s^T\end{bmatrix}$$

Ilustrasi dekomposisi matriks dengan TSVD

(Anton & Rorres, 2013), telah diolah kembali

Melalui metode TSVD, A direpresentasikan dengan matriks $\tilde{A} = \tilde{\Sigma}\tilde{V}^T\ berukuran\ s \times n.$
$\tilde{A}$ inilah yang akan menjadi *input* FCM (Burden *et al.*, 2011).

Metode Pendeteksian Topik: EFCM sebagai Metode *Clustering*

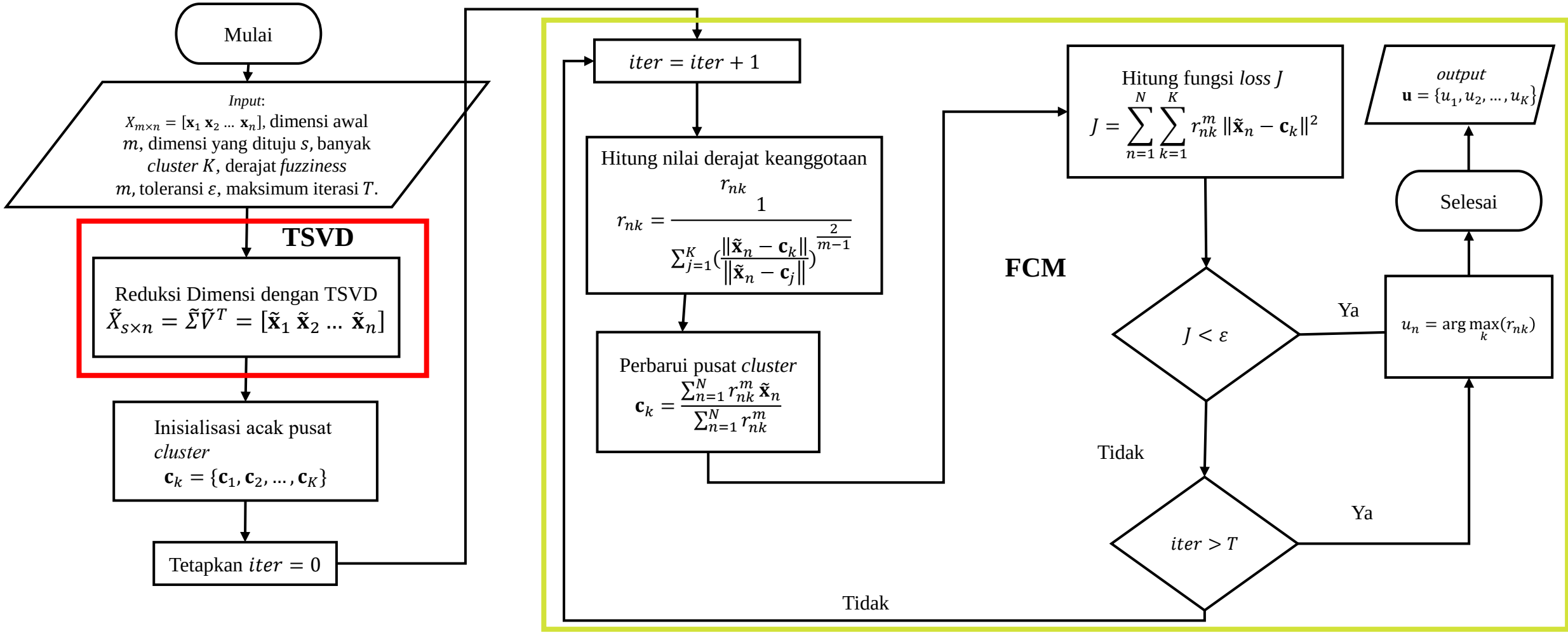
FCM

Fuzzy C-Means (FCM), salah satu teknik *soft clustering*, mengelompokkan data dengan mempertimbangkan tingkat keanggotaannya dalam setiap *cluster* (Bezdek *et al.*, 1984).

FCM memiliki fungsi objektif jarak *J*, yang mengukur jarak antara setiap data dengan *centroid*.

Fungsi tersebut memiliki parameter $r_{nk}^{m} \in [0,1]$ yang mengindikasikan tingkat keanggotaan data ke-*n* dalam *cluster* ke-*k* dan derajat *fuzzy* $m > 1$.

Metode Pendeteksian Topik: EFCM sebagai Metode *Clustering*



Metode Pendeteksian Topik: GPT sebagai Interpretasi Topik

Class-based Term Frequency Inverse Document Frequency (c-TFIDF)

c-TFIDF digunakan untuk **membuat kumpulan kata yang merepresentasikan topik** dengan memanfaatkan prinsip dari TFIDF untuk suatu dokumen dalam kelompok data yang telah dilakukan *cluster* (Grootendorst, 2022).

$$w_{t,c} = tf_{t,c} \times \log\left(1 + \frac{A}{tf_t}\right)$$

Keterangan:

$w_{t,c}$ bobot kata t pada kelas c,

$tf_{t,c}$ banyak kata t pada kelas c,

A rata-rata jumlah kata per kelas,

tf_t frekuensi kata dari seluruh kelas.

Metode Pendeteksian Topik: GPT sebagai Interpretasi Topik

Input berupa **teks** → Model **GPT** → **Topik yang Koheren**

- ***Prompt*** (instruksi atau pertanyaan) untuk tugas interpretasi topik
- **[KEYWORDS]** adalah ***topic words*** yang diperoleh dari c-TFIDF dan **[DOCUMENTS]** adalah **dokumen-dokumen representatif** untuk merepresentasikan setiap *topic words*. (Grootendorst, 2023)

Metode Pendeteksian Topik: GPT sebagai Interpretasi Topik

Contoh *Prompt* untuk Tugas Intrepretasi Topik

Tentukan topik utama dari [KEYWORDS] berdasarkan [DOCUMENTS] yang diberikan. Respon hanya dengan satu kalimat. Gunakan Bahasa Indonesia yang formal.

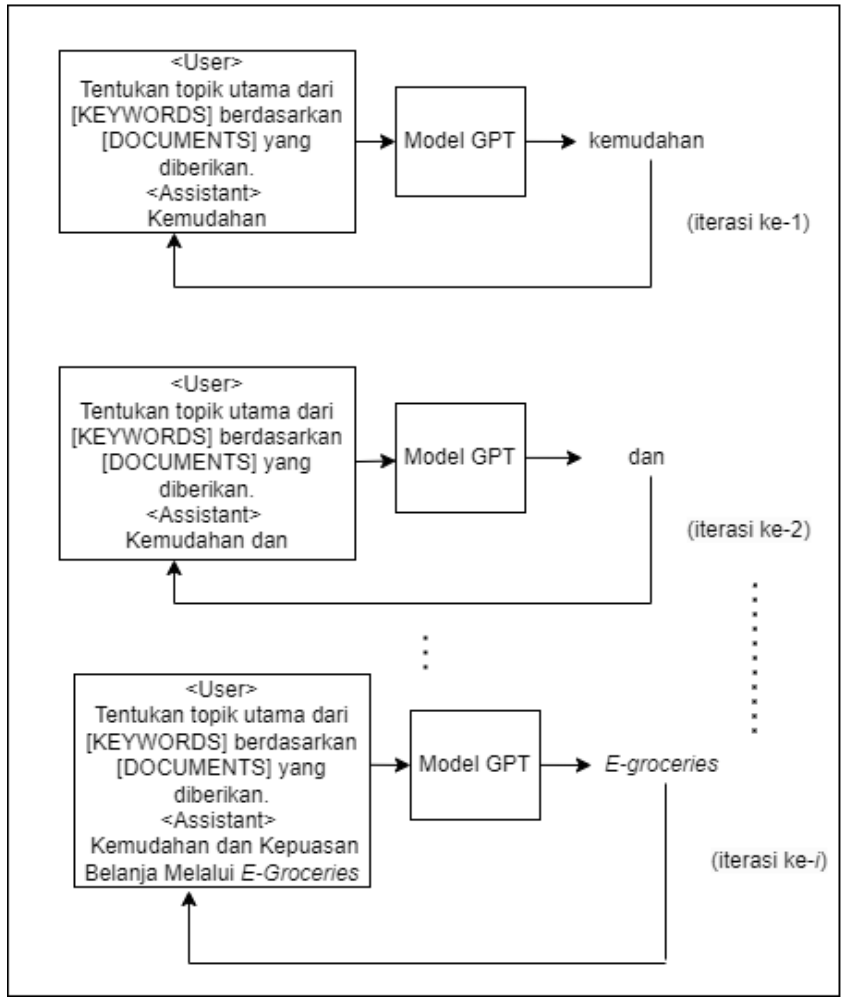
Contoh [KEYWORDS] dan [DOCUMENTS]

[KEYWORDS]: *e-groceries*, *online*, kenyamanan, pengiriman, pelanggan

[DOCUMENTS]:

- *E-groceries* memudahkan belanja *online* dengan pengiriman ke rumah.
- Pelanggan puas dengan produk beragam dan deskripsi rinci.
- Pengiriman tepat waktu dan barang segar sangat penting.
- *E-groceries* hemat waktu dengan pelacakan dan jadwal fleksibel.

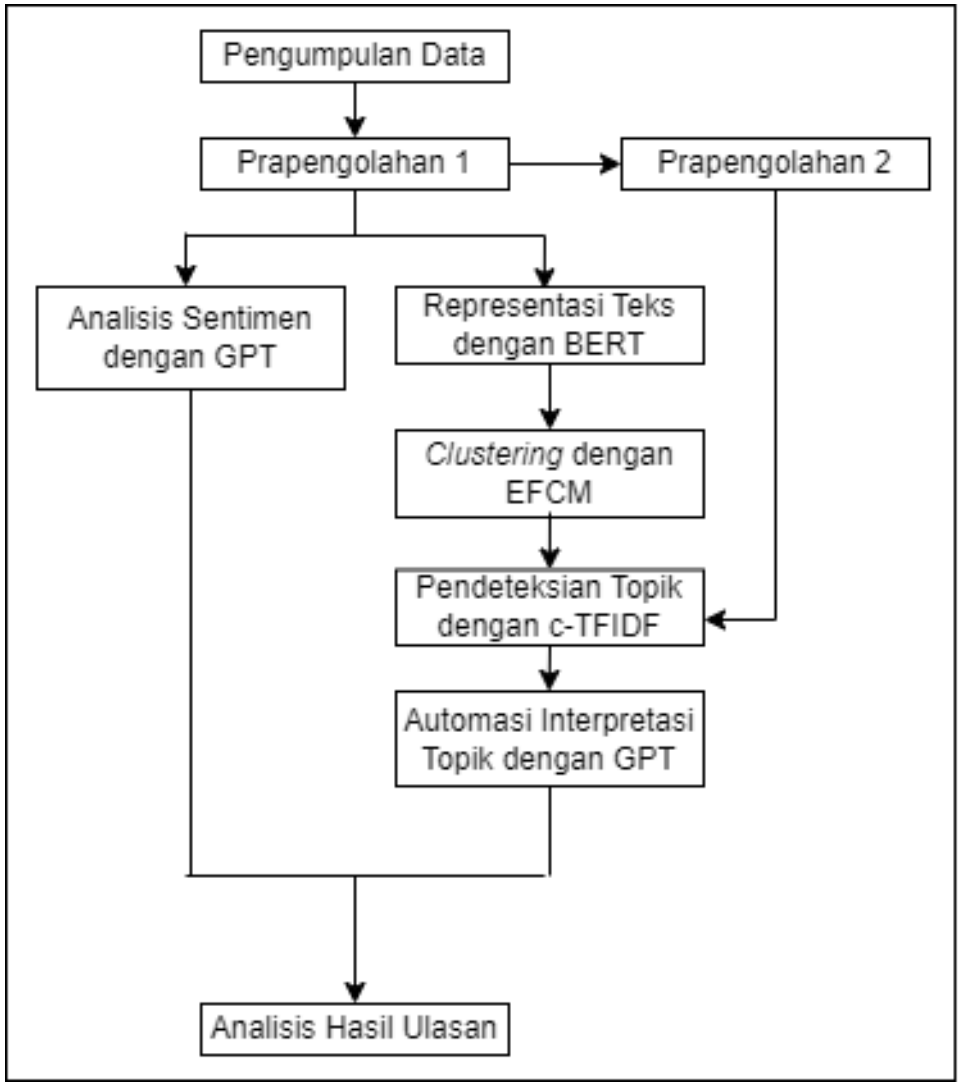
Metode Pendeteksian Topik: GPT sebagai Interpretasi Topik



Simulasi dan Analisis Hasil

Data
Penerapan Analisis Sentimen
Penerapan Pendeteksian Topik
Gabungan Hasil Analisis Sentimen dan Pendeteksian Topik
Analisis Rekomendasi untuk Perusahaan *E-groceries*

Tahapan Umum Simulasi



Data:
Pengumpulan Data

Pengambilan data dilakukan menggunakan *package* **"google-play-scraper"** pada Python Google Colaboratory.

Diperoleh **5.254** ulasan dengan **Bahasa Indonesia** paling relevan, berdasarkan Play Store, pada rentang waktu **31 Desember 2022 – 31 Desember 2023**.

Data: Prapengolahan 1

Sebelum Prapengolahan Data	Sayur dan buah yang dikirim segar, tapi pengiriman agak lama. Semoga next order bisa lebih cepat 😊
Setelah Prapengolahan 1	sayur dan buah yang dikirim segar, tapi pengiriman agak lama. semoga next order bisa lebih cepat

BERT optimal pada data dengan **prapengolahan minimal** (Alzahrani & Jololian, 2021), begitu pula dengan **GPT** (Radford *et al*., 2019).

Tahapan Prapengolahan 1:

1. Konversi huruf kapital menjadi huruf kecil.
2. Eliminasi tagar, mention, dan URL.
3. Penghapusan emotikon.
4. Penghapusan spasi berlebih.
5. Penyaringan ulasan dengan kurang dari tiga kata.
6. Pembuangan ulasan yang tidak relevan dengan aplikasi.

Data: Prapengolahan 2

Sebelum Prapengolahan Data	Sayur dan buah yang dikirim segar, tapi pengiriman agak lama. Semoga next order bisa lebih cepat 😊
Setelah Prapengolahan 1	sayur dan buah yang dikirim segar, tapi pengiriman agak lama. semoga next order bisa lebih cepat
Setelah Prapengolahan 2	sayur buah dikirim segar pengiriman lama order cepat

c-TFIDF menampilkan dokumen berdasarkan **frekuensi kemunculan kata**.

Tahapan Prapengolahan 2:

1. Penghapusan tanda baca.
2. Eliminasi angka.
3. Penghapusan *stopword*.
4. Penghapusan spasi berlebih.
5. Penyaringan ulasan dengan kurang dari 3 kata.

Jumlah ulasan:

- **Sebelum** prapengolahan data : **5.254**
- **Setelah** prapengolahan data : **3.078**

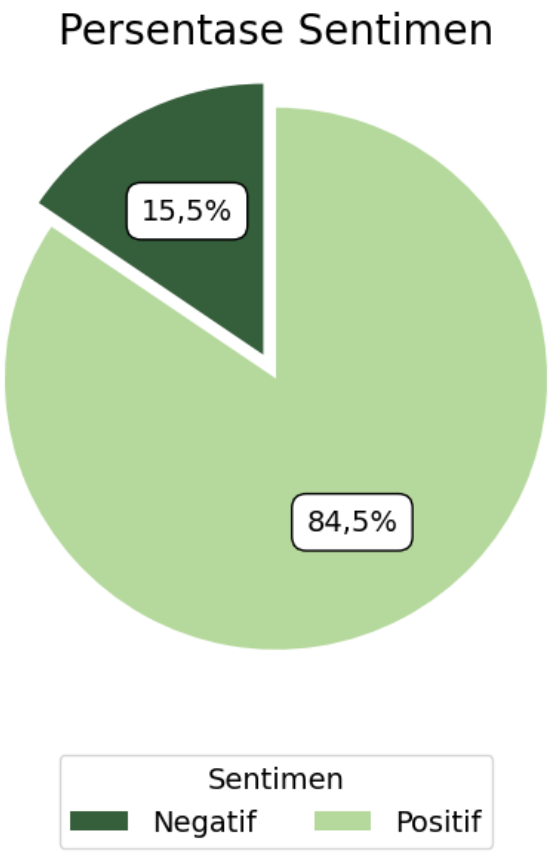
Penerapan Analisis Sentimen: Implementasi Model GPT

Hyperparameter

Hyperparameter	**Argumen**
Model yang digunakan (*model*)	gpt-3.5-turbo
Nilai *randomness* (*temperature*)	0

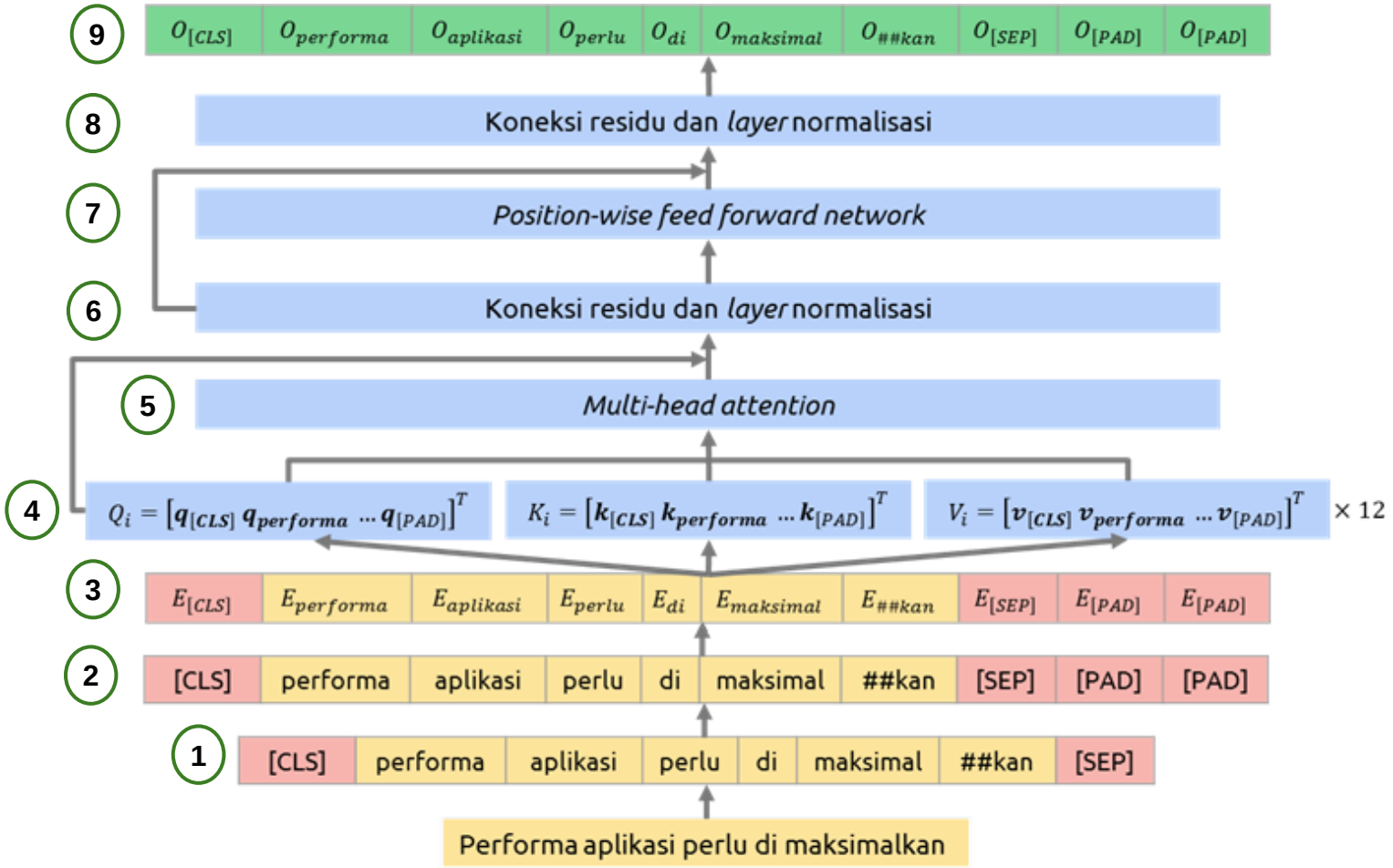
(Kheiri & Karimi, 2023)

Penerapan Analisis Sentimen: Hasil Analisis Sentimen



Setelah dilakukan klasifikasi sentimen pada keseluruhan **3.078 ulasan**, didapatkan 2.600 (**84,5%**) ulasan dengan sentimen **positif** dan 478 (**15,5%**) ulasan dengan sentimen **negatif**.

Penerapan Pendeteksian Topik: Representasi *Input* Menggunakan BERT



Ilustrasi representasi teks dengan BERT (Syamsyuriani, 2021)

1. **Tokenisasi** *input* dan penambahan token khusus.
2. Menyamakan panjang dokumen dengan ***padding***.
3. Menentukan representasi input (***embedding***).
4. Menentukan matriks ***query***, ***key***, dan ***value***, Q_i, K_i, V_i.
5. Mengoperasikan ***multi-head self attention***.
6. Menerapkan **koneksi residu** dan **normalisasi**.
7. Menerapkan ***position-wise feed forward network***.
8. Menerapkan **koneksi residu** dan **normalisasi**.
9. Diperoleh ***output*** yang akan menjadi ***input*** untuk lapisan berikutnya.

Penerapan Pendeteksian Topik: Implementasi EFCM sebagai Metode *Clustering*

Hyperparameter

Hyperparameter	**Argumen**
Banyak komponen utama TSVD (*n_components*)	5
Derajat *fuzzy* (*m*)	1,1
Ambang batas toleransi (*error*)	10^{-4}
Iterasi maksimal (*maxiter*)	200
Inisialisasi matriks *fuzzy c-partitioned* (*init*)	None

(Subakti *et al.*, 2022)

Penerapan Pendeteksian Topik: Pencarian Jumlah Topik Terbaik

$$TC - W2V = \frac{1}{\binom{N}{2}} \sum_{j=2}^{N} \sum_{i=1}^{j-1} similarity(t_j, t_i)$$

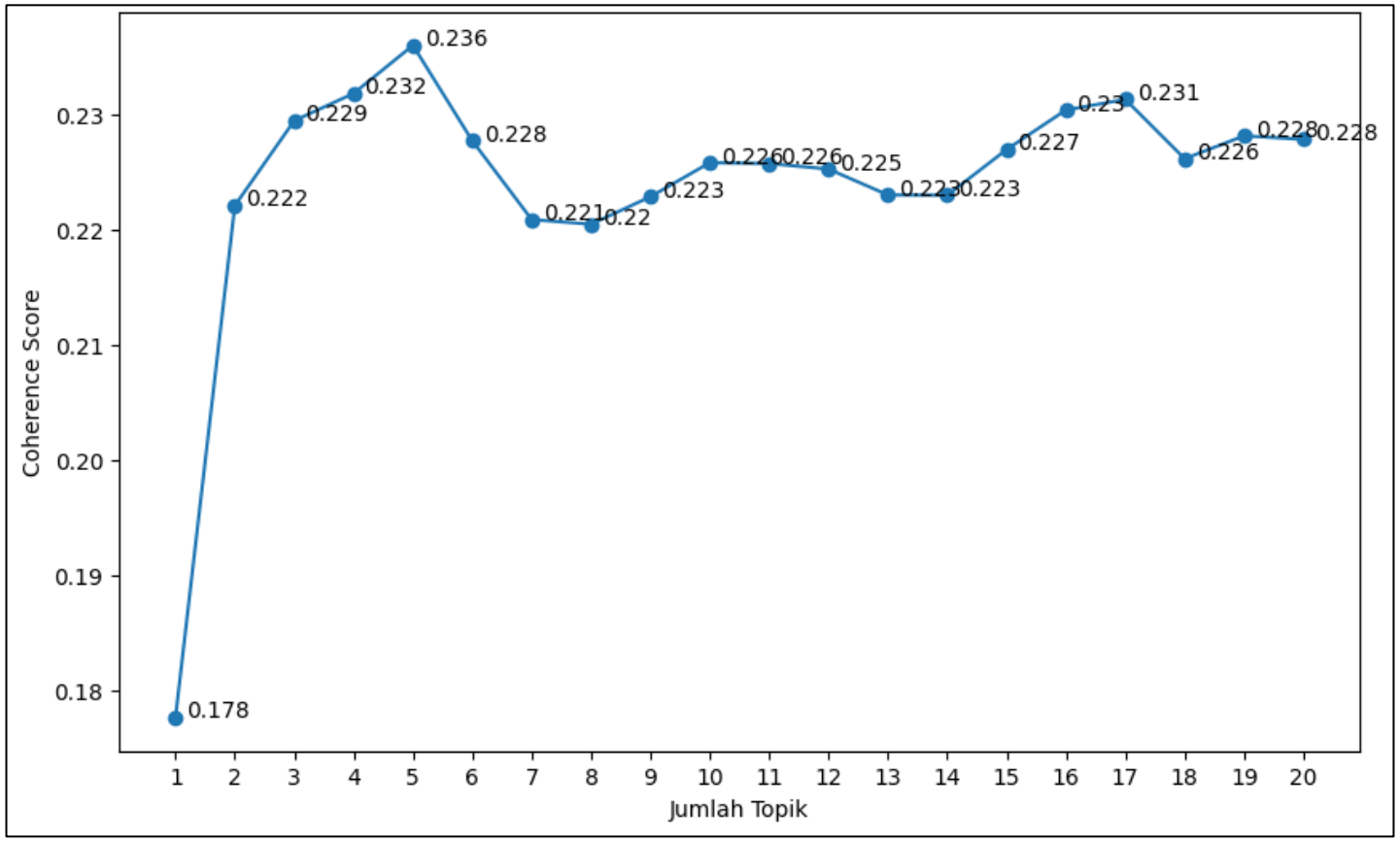
Keterangan:

- N banyak fitur *key words* pada suatu topik,
- $similarity(t_j, t_i)$ similaritas kosinus kata t_j dan t_i

Membantu menentukan jumlah topik optimal yang akan digunakan dalam pemodelan.

Mengevaluasi model berdasarkan tingkat koherensi topik menggunakan *Topic Coherence* Word2Vec (TC-W2V) (O'callaghan, 2015)

Penerapan Pendeteksian Topik: Pencarian Jumlah Topik Terbaik



Hasil *coherence* antar jumlah topik

Tahapan Pencarian Topik Terbaik:

1. Dicari jumlah topik terbaik antara 1 sampai 20 dengan membandingkan nilai koherensi menggunakan metode *Topic Coherence* Word2Vec (Murfi, 2021).
2. Dipilih jumlah topik yang memiliki rata-rata nilai *coherence* paling tinggi.

Hasil:

- Nilai rata-rata *coherence* tertinggi: **0,236**
- Jumlah topik: **5**

Penerapan Pendeteksian Topik: Implementasi Interpretasi Topik dengan Model GPT

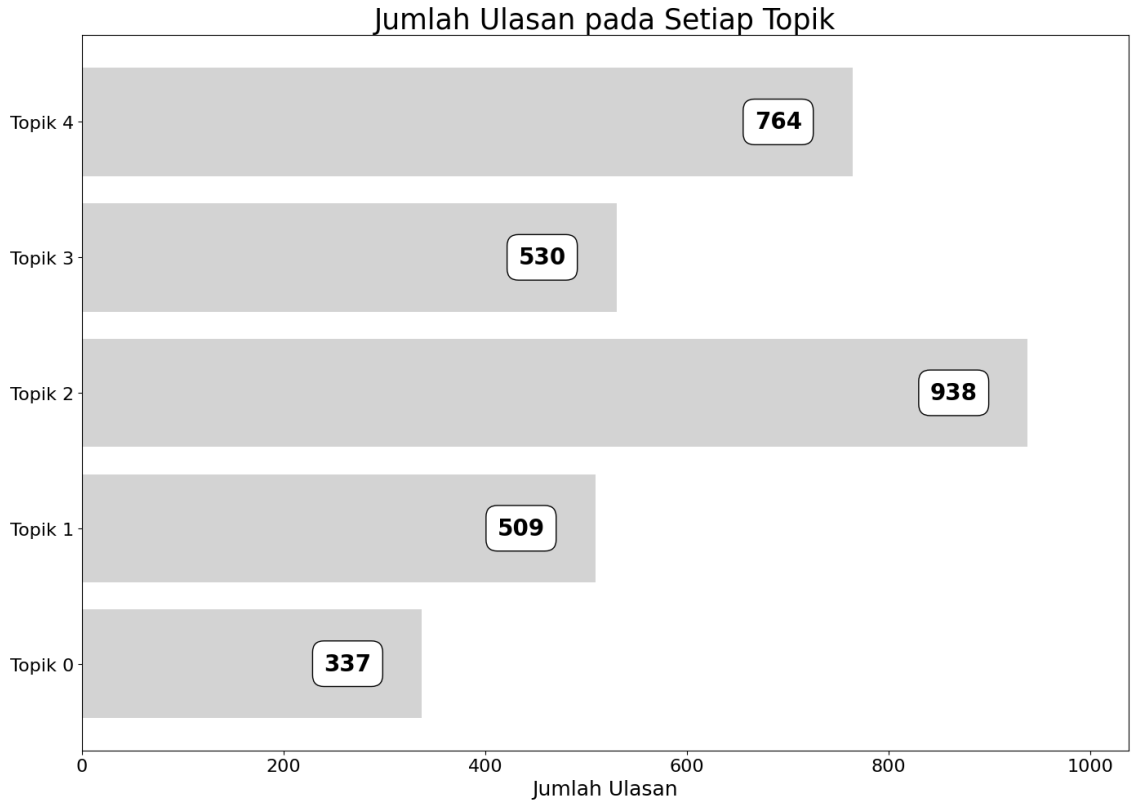
Hyperparameter

Hyperparameter	**Argumen**
Model (*model*)	gpt-3.5-turbo
Banyak dokumen terkait dengan kata kunci (*nr_docs*)	4
Waktu penundaan respon (*delay_in_seconds*)	*None*

(Grootendorst, 2023)

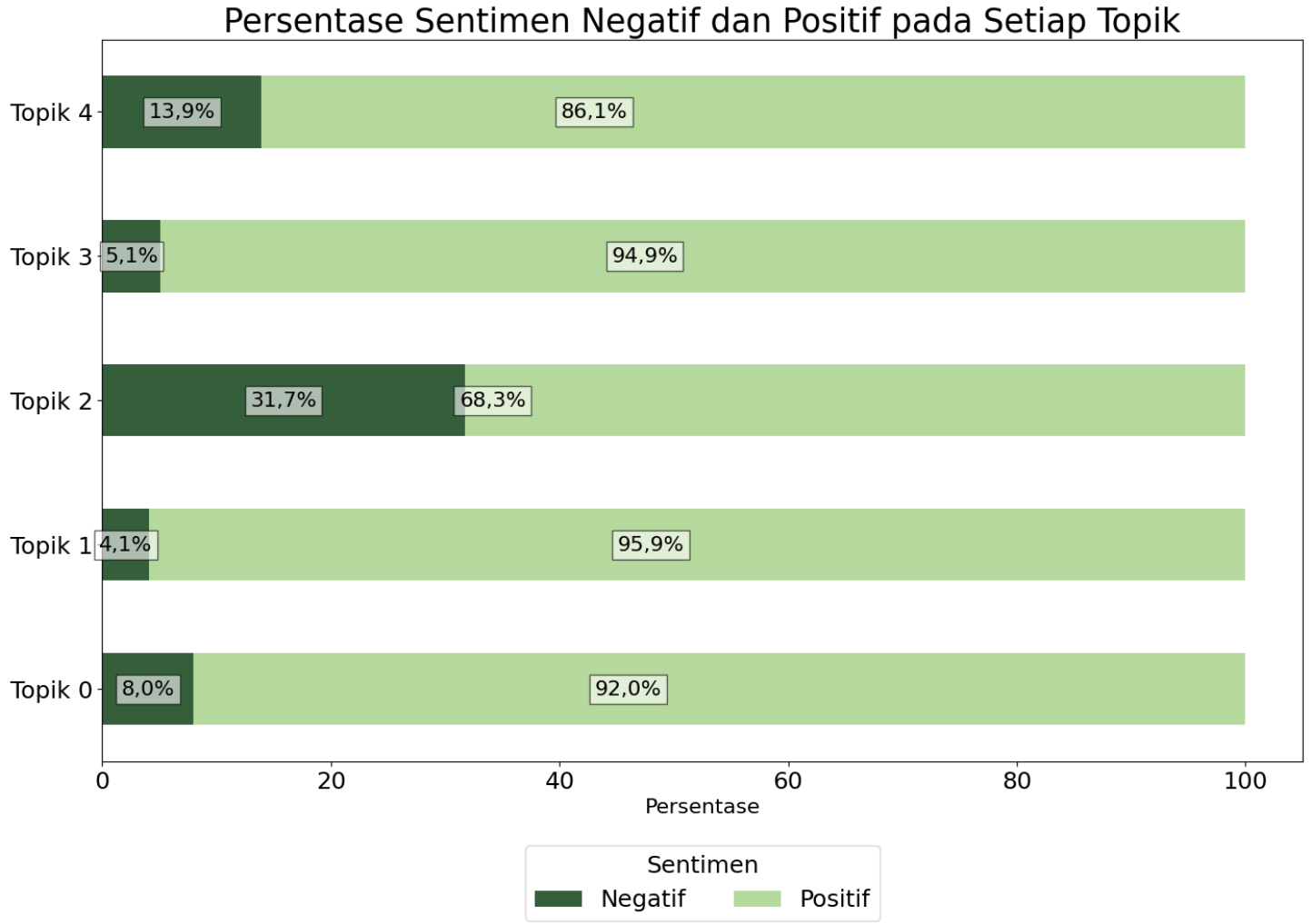
Penerapan Pendeteksian Topik: Hasil Pendeteksian Topik

Label Topik	Topik
Topik 1	Kecepatan dalam proses pengiriman barang yang bagus, segar, dan *fresh* dengan harga terjangkau, serta kemasan yang baik.
Topik 2	Pengiriman cepat sayur buah segar yang bagus.
Topik 3	Pengalaman berbelanja dan kesan pelanggan terhadap layanan Segari.
Topik 4	Kualitas produk yang segar dan bagus.
Topik 5	Kepuasan berbelanja dengan produk yang segar.



Distribusi topik berdasarkan jumlah ulasan

Gabungan Hasil Analisis Sentimen dan Pendeteksian Topik



Analisis Rekomendasi untuk Perusahaan *E-groceries*

Topik yang memiliki **persentase sentimen negatif** paling **signifikan** mengindikasikan **perlu** adanya **perbaikan**.

→

Topik 3

→

- Peningkatan yang *user-friendly* aplikasi
- Fitur *wishlist* dan notifikasi
- Peningkatan responsivitas layanan pelanggan
- Implementasi *chatbot* berbasis AI

Penutup

Kesimpulan
Saran

Kesimpulan

1. Berdasarkan hasil pendeteksian topik yang dilakukan, diperoleh lima topik utama. Topik-topik utama yang dibicarakan oleh pengguna mencakup kecepatan pengiriman barang, kualitas produk yang segar, harga yang terjangkau, kemasan yang baik, pengalaman berbelanja, serta kepuasan berbelanja dengan produk yang segar.

2. Berdasarkan analisis sentimen pada level topik yang dilakukan, diperoleh sentimen pada setiap topik dominan positif terhadap aplikasi Segari. Terdapat satu topik yang memiliki sentimen negatif yang signifikan, yakni topik mengenai pengalaman berbelanja dan kesan pelanggan terhadap layanan dengan persentase sentimen negatif 31,7%.

3. Berdasarkan gabungan hasil analisis sentimen dan pendeteksian topik, diberikan rekomendasi untuk topik yang memiliki sentimen negatif yang signifikan. Rekomendasi yang diberikan berdasarkan analisis dan pendapat pribadi penulis. Segari dapat meningkatkan pengalaman pengguna dengan membuat aplikasinya lebih intuitif dan *user-friendly*. Fitur seperti *wishlist*, notifikasi ketersediaan produk, dan riwayat pembelian dapat meningkatkan kenyamanan berbelanja. Penerapan *chatbot* berbasis AI dapat membantu memberikan solusi cepat dan efisien bagi pelanggan. Selain itu, pelatihan intensif bagi staf *customer service* dapat diberikan untuk menangani keluhan atau pertanyaan dengan cepat dan tepat. Rekomendasi yang diberikan memerlukan studi lebih lanjut dan peran ahli apabila akan diimplementasikan.

Saran

1. Mengembangkan penelitian dengan menganalisis ulasan berdasarkan kategori produk yang lebih spesifik seperti sayur, buah, dan daging.

2. Menganalisis pengaruh pembaruan fitur atau perubahan kebijakan pada aplikasi Segari terhadap sentimen pengguna dan topik yang dibahas.

3. Mengeksplorasi penggunaan *Large Language Model* (LLM) alternatif seperti Claude, Chinchilla, Gemini, atau Bloom sebagai pembanding GPT dalam analisis sentimen dan pendeteksian topik pada ulasan aplikasi *e-groceries*.

Daftar Pustaka

A'la, F. Y. (2022). Indonesian Sentiment Analysis towards MyPertamina Application Reviews by Utilizing Machine Learning Algorithms. *Journal of Informatics, Information System, Software Engineering and Applications*, 5(1), 80–91. https://doi.org/10.20895/INISTA.V5I1.838.

Aggarwal, C. C. (2018). Neural Networks and Deep Learning. *Springer International Publishing*. https://doi.org/10.1007/978-3-319-94463-0.

Ainslie, J., Lee-Thorp, J., de Jong, M., Zemlyanskiy, Y., Lebrón, F., & Sanghai, S. (2023). GQA: *Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints*. arXiv preprint arXiv:2305.13245.

Alatas, H., Murfi, H., & Bustamam, A. (2018). Topic Detection using fuzzy c-means with nonnegative double singular value decomposition initialization. *Int. J. Advance Soft Compu. Appl*, 10(2).

Alghamdi, R., & Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. (IJACSA) *International Journal of Advanced Computer Science and Applications*, 6(1), 147–153. www.ijacsa.thesai.org.

Anton, H., & Rorres, C. (2013). Elementary Linear Algebra: Applications Version (11th ed.). John Wiley & Sons.

Barde, B. V., & Bainwad, A. M. (2017). An Overview of Topic Modeling Methods and Tools. *International Conference on Intelligent Computing and Control Systems* (ICICCS), 745–750.

Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic Topic Models: A focus on graphical model design and applications to document and image analysis. *IEEE Signal Processing Magazine*, 27(6), 55–65. https://doi.org/10.1109/MSP.2010.938079.

Daftar Pustaka

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models Are Few-Shot Learners. *Proceedings of the 34th International Conference on Neural Information Processing Systems*.

Chen, X., Chen, X., Ding, C. G., Ding, L., & Han, F. (2021). GPS-IMU fused navigation scheme for transport logistics applications. *IEEE Transactions on Transportation Electrification*, 7(4), 2114-2126.

Cichocki, A., & Phan, A. H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E92-A (3), 708–721. https://doi.org/10.1587/transfun.E92.A.708.

da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & Alves, S. F. dos R. (2017). Artificial Neural Networks: A Practical Course. Springer Cham. https://doi.org/10.1007/978-3-319-43162-8.

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2023). Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 512–515.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 1–16. https://doi.org/10.48550/arXiv.1810.04805.

e-Conomy SEA (2023). e-Conomy SEA 2023 report. https://economysea.withgoogle.com/report/

Daftar Pustaka

Févotte, C., & Idier, J. (2010). Algorithms for nonnegative matrix factorization with the beta-divergence. http://arxiv.org/abs/1010.1763

Garcia, K., & Berton, L. (2021). Topic Detection and Sentiment Analysis in Twitter Content Related to COVID-19 from Brazil and the USA. *Applied Soft Computing*, 101. https://doi.org/10.1016/j.asoc.2020.107057.

Gatta, V., Marcucci, E., & Le Pira, M. (2023). 21. E-commerce and urban logistics: trends, challenges, and opportunities. *Handbook on City Logistics and Urban Freight*: 0, 422.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Grootendorst, M. (2023). Topic modeling with Llama 2. Medium. https://towardsdatascience.com/topic-modeling-with-llama-2-85177d01e174

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). “I think this is the most disruptive technology”: Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. http://arxiv.org/abs/2212.05856

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online Learning for Latent Dirichlet Allocation.

Jagani, K., Oza, F. V., & Chauhan, H. (2020). Customer Segmentation and Factors Affecting Willingness to Order Private Label Brands: An E-Grocery Shopper's Perspective. *In Improving Marketing Strategies for Private Label Products* (pp. 227-253). IGI Global.

Jamal, U. (2023). pytorch-llama.github.com.

Daftar Pustaka

Kheiri, K., & Karimi, H. (2023). SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning. https://arxiv.org/pdf/2307.10234v2.pdf

Kumar, V. M., Prasad, K., & Gopala, M. (2020). Cold Chain Technology for Agricultural Produce: Socio-Economic and Environmental Implications. *Journal of Agricultural and Environmental Ethics*, 33(5), 717-735.

Kung, J., O'Donnell, M., & Ellison, N. (2023). GPT-3: Language Models are Few-Shot Learners. *Journal of Machine Learning Research*, 23(124), 1-50.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publisher.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35.

Maalej, W., & Nabil, H. (2015, August). Bug report, feature request, or simply praise? on automatically classifying app reviews. *In 2015 IEEE 23rd international requirements engineering conference* (RE) (pp. 116-125). IEEE.

Mishra, A., Nautiyal, S., & Dhabaleswar, M. (2022). Analysis of Customer Satisfaction through e-Groceries Apps Post COVID-19: An Empirical Study. *Journal of Retailing and Consumer Services*, 65, 1-10.

Muliawati, T., & Murfi, H. (2017). Eigenspace-based Fuzzy C-Means for Sensing Trending Topics in Twitter. *AIP Conference Proceedings*, 1862. https://doi.org/10.1063/1.4991244.

Daftar Pustaka

Murfi, H. (2021). A scalable eigenspace-based fuzzy c-means for topic detection. *Data Technologies and Applications*, 55(4), 527–541. https://doi.org/10.1108/DTA-11-2020-0262.

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.

Nicholas Ramos Richardo. (2022). Analisis Performa EFCM dengan BERT sebagai Representasi Teks pada Pendeteksian Topik. Universitas Indonesia.

Nguyen, T. H., Xuan, T. H., & Hoang, D. T. (2020). The Efficacy of RoBERTa Model for Named Entity Recognition. *Journal of Artificial Intelligence Research*, 68, 343-354.

Nguyen, T. H., Xuan, T. H., Hoang, D. T., & Le, Q. V. (2022). Pricing Strategies in E-commerce: An Empirical Study. *Journal of Retailing and Consumer Services*, 65, 102118.

Papadopoulos, S., Corney, D., & Maria Aiello, L. (2014). SNOW 2014 Data Challenge: Assessing the Performance of News Topic Detection Methods in Social Media. *Proceedings of the SNOW 2014 Data Challenge*, 1–8. http://ceur-ws.org.

Play Store. (2023). Play Store. https://play.google.com/store/apps

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. https://api.semanticscholar.org/CorpusID:160025533.

Daftar Pustaka

Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. *In Proceedings of the first instructional conference on machine learning* (Vol. 242, No. 1, pp. 29-48).

Ravichandiran, S. (2021). Getting Started with Google BERT: Build and train state-of-the-art natural language processing models using BERT. *Packt Publishing Ltd*.

Shazeer, N. (2020). Glu variants improve transformer. arXiv preprint arXiv:2002.05202.

Su, J., Lu, Y., Pan, S., Wen, B., & Liu, Y. (2021). RoFormer: Enhanced Transformer with Rotary Position Embedding. ArXiv, abs/2104.09864. https://api.semanticscholar.org/CorpusID:233307138.

Subakti, A., Murfi, H., & Hariadi, N. (2022). The performance of BERT as data representation of text clustering. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-022-00564-9

Usharani, M. (2018). Sentiment Analysis using supervised learning algorithms. *International Journal of Engineering and Advanced Technology*, 8(2), 167-172.

Uthirapathy, S. E., & Sandanam, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *Procedia Computer Science*, 218, 908–917. https://doi.org/10.1016/j.procs.2023.01.071.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*. https://api.semanticscholar.org/CorpusID:13756489.

Daftar Pustaka

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the* ACM, 9, 36 – 45.

Wijaya, D., Murfi, H., & Ardaneswari, G. (2024). Topic-level sentiment analysis for user reviews in gasoline subsidy application [Accepted]. *The 11th IEEE Swiss Conference on Data Science, Switzerland*.

Williams, P., & Naumann, E. (2011). Customer satisfaction, retention, and loyalty: An empirical assessment of 19 years of research in business-to-business services. *Journal of Business-to-Business Marketing*, 19(7), 79-128.

Yudhistira Jinawi Agung. (2023). Analisis Sensitivitas Parameter Model EFCM Berbasis BERT untuk Pendeteksian Topik. Universitas Indonesia.

Yusdiansyah, M. R., Murfi, H., & Wibowo, A. (2019). Randomspace-Based Fuzzy C-Means for Topic Detection on Indonesia Online News. In R. Chamchong & K. W. Wong (Eds.), *Multi-disciplinary Trends in Artificial Intelligence* (pp. 133–143). Springer International Publishing.

Zhang, B., & Sennrich, R. (2019). Root mean square layer normalization. *Advances in Neural Information Processing Systems*, 32.

Zhang, B., Yang, H., Zhou, T., Babar, A., & Liu, X. Y. (2023). *Enhancing financial sentiment analysis via retrieval augmented large language models*. arXiv preprint arXiv:2310.04027.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

Terima Kasih

TSVD

Misal *A output* dari BERT berukuran *n* × *768,* dan banyak komponen utama TSVD=5:

$$\tilde{A} = A^T$$

$$\tilde{A}_{768\times n} \approx \tilde{U}_{768\times 5}\tilde{\Sigma}_{5\times 5}(\tilde{V}_{n\times 5})^T \approx [u_1\ u_2 \dots\ u_5]\begin{bmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_5\end{bmatrix}\begin{bmatrix}v_1^T\\ v_2^T\\ \vdots\\ v_5^T\end{bmatrix}$$

Dengan $\tilde{U}$ adalah matriks ortogonal berukuran 768×5, $\tilde{\Sigma}$ adalah matriks berukuran 5×5, dan $\tilde{V}^T$ merupakan transpos dari matriks $\tilde{V}$ berukuran $5\times n$.

$$\tilde{A}_{5\times n} = \tilde{\Sigma}_{5\times 5}\left(\tilde{V}_{n\times 5}\right)^T$$

Maka *X input* dari EFCM:

$$X_{n\times 5} = (\tilde{A}_{5\times n})^T$$

BPE

Konsep Dasar BPE:

1. Mulai dengan karakter individual.
2. Temukan pasangan karakter yang sering muncul
3. Gantikan dengan simbol baru.
4. Ulangi proses hingga batas tertentu.

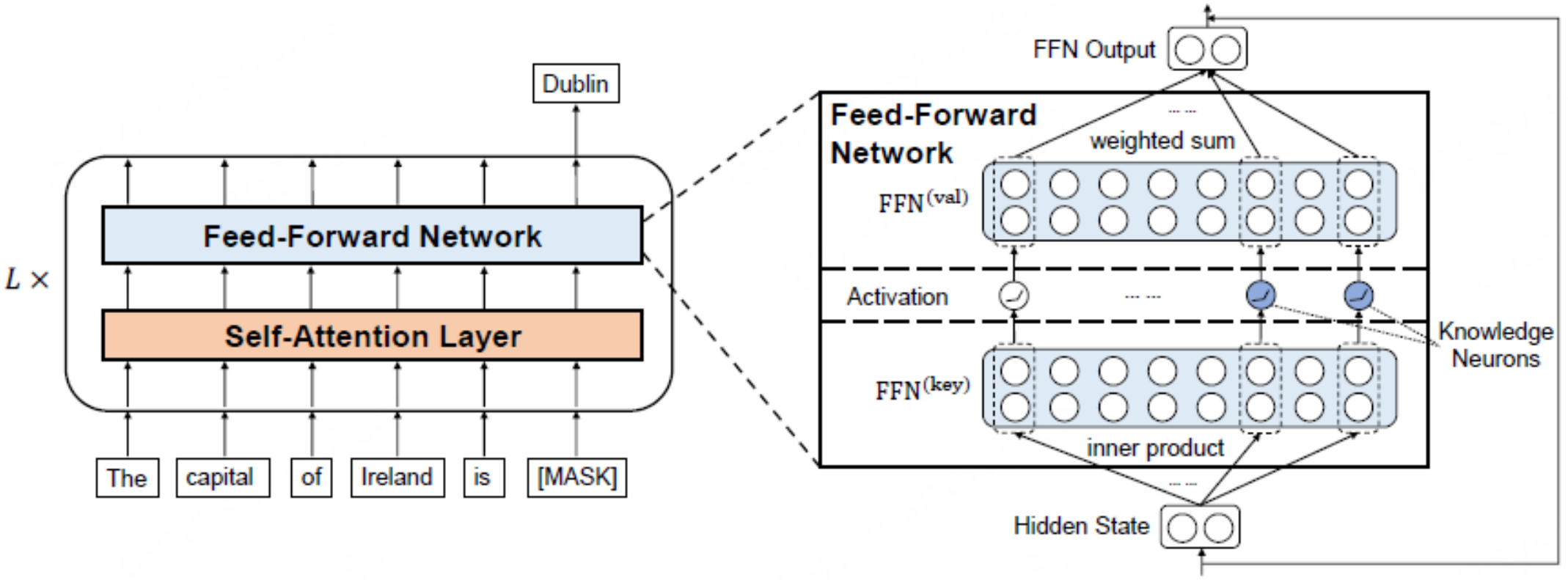
Contoh Pengaplikasian pada Kata "aplikasi":

1. Iterasi 1
   - Pasangan `a p` paling sering muncul.
   - Ganti `a p` dengan `ap`.
   - Hasil: `ap l i k a s i`
2. **Iterasi 2:**
   - Pasangan `li` paling sering muncul.
   - Ganti `li` dengan `li`.
   - Hasil: `ap li ka s i`
3. **Iterasi 3:**
   - Pasangan `i k` paling sering muncul.
   - Ganti `i k` dengan `ik`.
   - Hasil: `ap l ik asi`

Contoh Pasangan Kata dalam Korpus BPE:

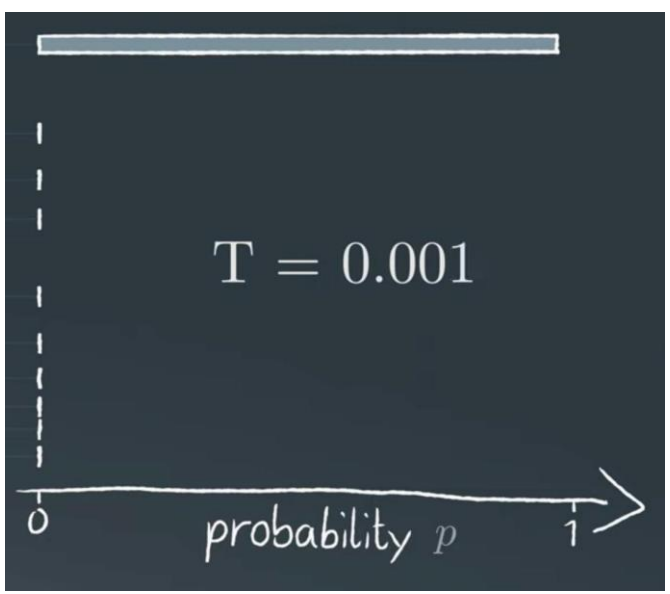
- ` a` - sub-kata yang sering muncul.
- ` in` - umum dalam kata seperti "informasi."
- ` ap`, `lik` - ditemukan dalam "aplikasi."

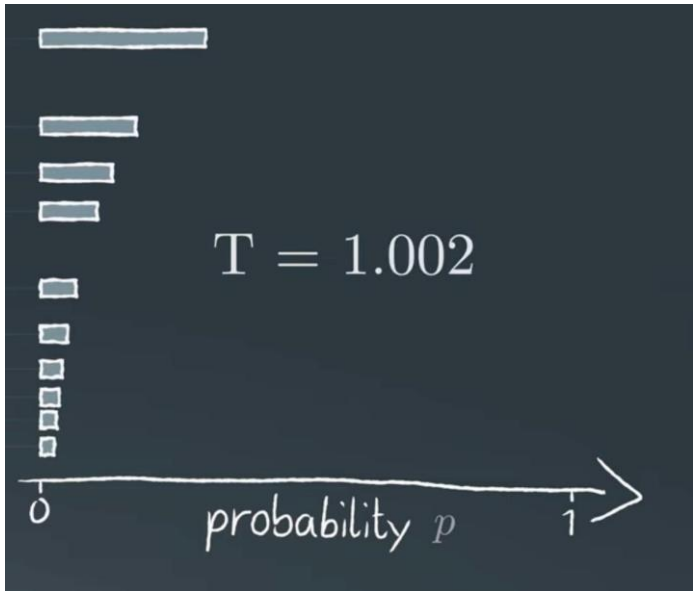
Ilustrasi FFN

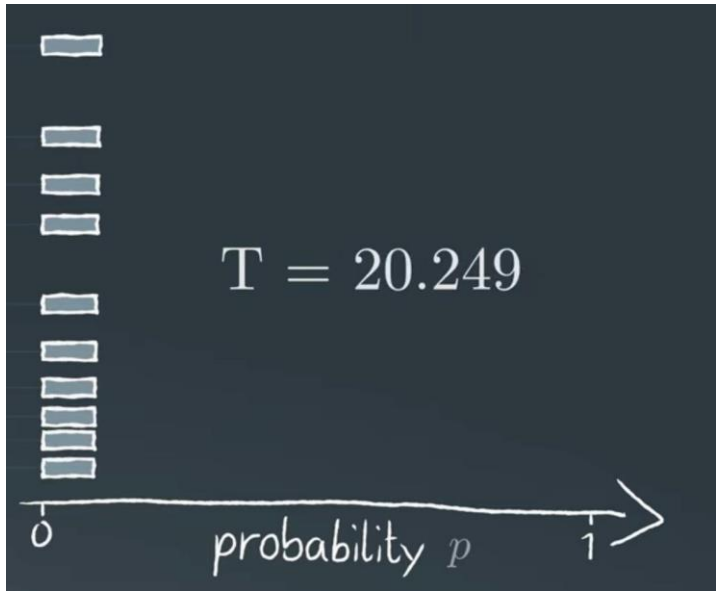


Temperature

$$P(x_i) = \frac{\exp\left(\frac{\log P'(x_i)}{T}\right)}{\sum_j \exp\left(\frac{\log P'(x_j)}{T}\right)}$$







Marble, 2023

Dengan *temperature*=0, hasilnya konsisten. Setiap kali teks ulasan yang sama diberikan ke model dengan pengaturan ini, hasil yang dihasilkan akan selalu "Negatif," memastikan keandalan dalam prediksi sentimen tanpa adanya fluktuasi atau variasi.