BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini merupakan bahasan terkait dari hasil serta pembahasan yang bertujuan untuk menjawab rumusan masalah penelitian. Setiap hasil penelitian yang di dapat bertujuan untuk menjawab rumusan masalah serta tujuan penelitian untuk memberikan pemahaman yang baik mengenai efektifitas serta kontribusi sistem yang dikembangkan.

4. 1. Hasil Serta Pembahasan Rumusan Masalah Penelitian 1

Dalam penelitian ini merupakan penelitian untuk melakukan analisis sentimen terhadap isu *"marriage is scary"* pada media sosial X. Proses analisis sentimen ini memanfaatkan proses pelatihan dengan menggunakan BERT (*Bidirectional Encoder Representations from Transformers*). Proses pelatihan model memanfaatkan jenis model *pretrained* BERT yakni *bert-base-multilangual-cased* dari *hugging face* Google. Model ini dipilih karena telah mampu dalam melakukan pengenalan pada 104 bahasa.

Batasan dan serta tahapan evaluasi dilakukan dalam penelitian ini adalah untuk melakukan analisis sentimen, serta mengetahui kinerja dari model tersebut dalam melakukan klasifikasi sentimen pada data dua negara yakni Indonesia dan Amerika Serikat. Maka dari itu, penelitian ini berfokus kepada evaluasi hasil kinerja model dan hasil klasifikasi analisis sentimen di dua negara tersebut.

4. 2. 1. Hasil Proses *Business Understanding*

Tahapan pertama dalam penelitian ini dimulai dengan pemahaman konteks bisnis atau latar belakang masalah yang ingin diselesaikan. Tujuan dari penelitian ini yakni untuk memahami mengapa isu ini penting untuk diteliti dan hasil seperti apa yang diharapkan. Dalam konteks ini, media sosial X (Twitter) dipilih karena merupakan platform yang banyak digunakan masyarakat untuk menyampaikan opini, sehingga sangat relevan untuk dilakukan analisis sentimen.

a) Identifikasi Tujuan Penelitian

Pada tahapan awal proses penelitian, dilakukan pemahaman terhadap konteks bisnis dan tujuan penelitan. Tujuan utama dari penelitian ini adalah untuk memahami opini publik terhadap isu *marriage is scary* melalui pendekatan analisis sentimen dengan algoritma *multilangual* BERT, serta memvisualisasikan tren

demografis yang muncul dari analisis data di dua negara, yakni Indonesia dan Amerika Serikat. Tujuan ini dicapai melalui beberapa langkah, yaitu:

a. Mengumpulkan data *tweet* yang berkaitan dengan topik "*Marriage is Scary*" menggunakan *Tweet-harvest*.
b. Melakukan *labeling* sentimen menjadi dua kategori: *positive* dan *negative*.
c. Melatih model menggunakan pendekatan *fine-tuning* terhadap model *Multilingual* BERT.
d. Menganalisis hasil model untuk memahami bagaimana persepsi masyarakat terhadap isu tersebut berbeda di tiap negara.

b) Kebutuhan dan Tantangan Penelitian

Dalam pelaksanaan penelitian, hasilnya ditemukan beberapa kebutuhan teknis dan tantangan utama yang memengaruhi keseluruhan proses penelitian. Salah satu kebutuhan utama dalam penelitian ini adalah memperoleh data yang relevan dan sesuai dengan isu *"Marriage is Scary"*. Berdasarkan hasil *scraping* menggunakan *Tweet-Harvest*, tidak semua *tweet* yang mengandung kata kunci tersebut memiliki konteks yang tepat. Oleh karena itu, proses pembersihan data dilakukan secara menyeluruh untuk *memfilter tweet* yang benar-benar menyampaikan opini terhadap pernikahan. Relevansi data sangat penting agar hasil analisis sentimen benar-benar mencerminkan opini publik, bukan hanya sekadar teks acak yang mengandung kata kunci.

Selain itu, proses pelatihan model BERT yang digunakan dalam penelitian ini membutuhkan daya komputasi yang cukup tinggi. Model *Multilingual BERT* termasuk dalam kategori *deep learning* dengan arsitektur kompleks, sehingga perangkat GPU menjadi kebutuhan penting untuk mempercepat proses *training*. Dalam pelaksanaannya, penelitian ini memanfaatkan fasilitas GPU gratis melalui *Google Colab* yang mampu mengurangi waktu pelatihan secara signifikan dibandingkan dengan hanya menggunakan CPU.

Salah satu kendala teknis yang cukup menonjol adalah ketidakseimbangan jumlah *tweet* dengan sentimen positif dan negatif. Pada data awal, jumlah *tweet* negatif terhadap isu pernikahan jauh lebih banyak dibandingkan yang positif pada *dataset* Indonesia. Ketidakseimbangan ini kemudian berpotensi membuat model

bias terhadap kelas mayoritas. Sehingga untuk mengatasi hal tersebut dilakukan dengan augmentasi data dengan *fill-mask* IndoBERT.

Tantangan lainnya adalah pada proses *scraping* data dihadapkan pada keterbatasan yang diberikan oleh pihak X (sebelumnya Twitter), terutama untuk akun pengembang dengan akses gratis. Selain itu juga, adanya batasan jumlah *scrolling* data dengan harus menunggu jeda waktu beberapa menit. Selain itu, adanya kebijakan API berbayar yang ditawarkan oleh X tergolong mahal dan tidak terjangkau untuk penelitian skala mahasiswa. Akibatnya, proses pengambilan data harus dilakukan strategi tertentu yaknik menggunakan *tweet-harvest* untuk tetap memperoleh data yang representatif.

c) Merancang Proyek Penelitian

Rencana proyek penelitian ini disusun secara sistematis agar bertujuan untuk memastikan setiap tahapan berjalan efektif dan efisien. Setiap fase dirancang berdasarkan kebutuhan metodologis, keterbatasan teknis, serta target penelitian. Dalam pelaksanaan penelitian, sebelumnya dilakukan perumusan beberapa hal utama seperti sumber daya yang digunakan, alat bantu analisis, metode pemrosesan data, serta output yang ingin dicapai. Untuk meminimalisir hambatan yang mungkin terjadi selama proses penelitian tersebut, kemudian dilakukan pula identifikasi risiko serta strategi untuk mengatasinya. Berikut merupakan ringkasan perencanaan proyek penelitian pada Tabel 4.1.

Tabel 4.1 Rancangan Proyek Penelitian

Komponen	Keterangan
Metodologi	CRISP-DM (*Cross Industry Standard Process for Data Mining*)
Sumber Data	*Dataset tweet* dari media sosial X (Twitter) dikumpulkan menggunakan *tweet-harvest* dengan kata kunci, rentang waktu, dan bahasa.
Tools & Library	*Python, Hugging face Transformers, TensorFlow/Keras, pandas, scikit-learn, matplotlib, seaborn*
Model Analisis	*Multilingual* BERT dengan *fine-tuning*
Output Penelitian	a. Model klasifikasi sentimen publik terhadap isu "*marriage is scary*" b. Visualisasi distribusi sentimen c. Perbandingan hasil antara Indonesia dan Amerika Serikat
Potensi Risiko	Data tidak seimbang, batasan keamanan pihak X, kebutuhan GPU untuk komputasi.

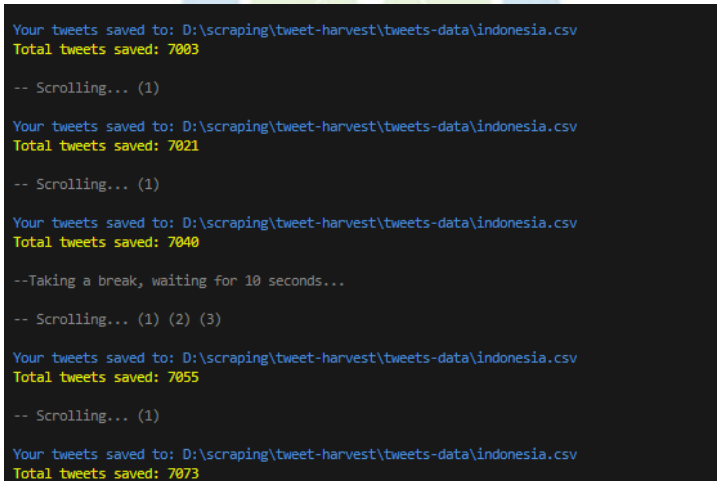
Komponen	Keterangan
Strategi Mitigasi	a. Mengunakan d*ata augmentation*. b. Maksimalkan penggunaan teknik *scraping.* c. Menggunakan *Google Colab* (*free GPU*)

4. 2. 2. Hasil *Data Understanding*

Hasil dari proses *data understanding* diketahui bahwa pada tahapan ini merupakan proses untuk mengenal dengan keadaan data penelitan. Proses yang dilakukan adalah proses untuk mendapatkan data yang relevan dengan penelitan, serta menganalisis data setelah proses pengambilan data.

a) Pengumpulan Data Penelitian

Pengumpulan data dalam penelitian ini dilakukan dengan teknik *scraping* menggunakan alat bantu bernama *tweet-harvest*. *Tools* tersebut merupakan sebuah *tool python* yang dapat mengambil data dari media sosial X (Twitter) melalui proses *scrolling* otomatis di laman *website* X. Metode ini dipilih karena lebih fleksibel dalam menangkap *tweet* yang relevan. Proses hasil *scraping data* pada media sosial X dapat dijelaskan pada Gambar 4.1 berikut.



Gambar 4.1 Proses *Scraping Dataset*

Gambar 4.1 merupakan proses *scraping* data dengan mengakses data menggunakan *auth_token*, dimana digunakan sebagai salah satu parameter utama agar proses *scraping* dapat dilakukan dengan lancar dan akun tetap dikenali oleh sistem. *Tweet* yang dikumpulkan berasal dari dua negara, yaitu Indonesia dan Amerika Serikat. Proses pengambilan data menggunakan filter kata kunci, rentang waktu, dan bahasa. Kata kunci yang digunakan untuk data Indonesia mencakup: *"marriage is scary", "takut nikah", "trauma nikah", "takut cerai",* dan

"pernikahan Indonesia". Sedangkan untuk data dari Amerika Serikat, kata kunci yang digunakan antara lain: *"marriage is scary", "fear of marriage", "boycott marriage", "marriage strike", "divorce",* dan *"be single"*. Adapun *filter* bahasa yang digunakan adalah *lang=id* untuk Indonesia dan *lang=en* untuk Amerika Serikat. Rentang waktu pengambilan data dimulai dari 1 Agustus 2024 hingga 20 Mei 2025, dikarenakan pada periode tersebut topik "*marriage is scary*" mulai menjadi perbincangan yang cukup aktif di media sosial X di Indonesia.

Setelah melalui proses pengambilan data pada media sosial X dengan menerapkan *tweet-harvest*, data ini akan digunakan untuk proses penelitian. Maka jumlah data yang telah terkumpul setelah proses *scrapping data* pada kedua negara dengan rentang dari tanggal 1 Agustus 2024 sampai 20 Mei 2025 dirincikan dalam Tabel 4.2 berikut.

Tabel 4.2 Jumlah *Dataset* yang Terkumpul

	Indonesia	Amerika Serikat
Jumlah *Tweet*	7.073	4.953

b) Menelaah *Dataset*

Setelah proses *scraping data,* hasil dari proses *scraping* ini kemudian disimpan dalam format CSV (*Comma-Separated Values*). Tujuan dari menyimpan dalam format ini adalah agar memudahkan dalam proses analisis data selanjutnya. File CSV yang telah tersimpan tersebut, untuk memisahkan antar bagiannya menggunakan separator koma (,). Selain itu *dataset* yang telah tersimpan terdiri dari sejumlah kolom penting yang mendukung analisis sentimen dan metadata. Adapun kolom-kolom yang dikumpulkan dalam *dataset* ini meliputi:

1) *conversation_id_str*: merupakan kolom data ID percakapan pada *tweet.*
2) *created_at*: merupakan kolom data berupa waktu dari *tweet* data diposting.
3) *favorite_count*: merupakan jumlah suka atau favorit pada *tweet.*
4) *full_text*: merupakan kolom berisi teks cuitan *tweet.*
5) *id_str*: ID unik berasal dari tiap *tweet.*
6) *image_url*: URL gambar yang yang ada dalam cuitan *tweet*.
7) *in_reply_to_screen_name*: menunjukkan *username* yang dibalas oleh *tweet* ini.

8) *lang*: merupakan kolom bahasa dari *tweet*. Untuk bahasa Indonesia maka data berisi *in* dan untuk bahasa Inggrias data berisi *en.*
9) *location*: berisi data lokasi dari profil pengguna.
10) *quote_count*: berisi data jumlah *quote retweet* dari *tweet.*
11) *reply_count*: jumlah balasan dari *tweet.*
12) *retweet_count*: jumlah *retweet* pada *tweet.*
13) *tweet_url*: URL atau tautan *tweet* pada platform X.
14) *user_id_str*: ID unik dari pengguna yang membuat *tweet*
15) *username*: berisi *username* dari pengguna yang membuat *tweet.*

Setiap baris dalam *dataset* tersebut mewakili satu *tweet* yang telah sesuai dengan hasil pencarian data, dimana *dataset* ini telah memiliki informasi lengkap mulai dari isi teks *tweet*, waktu unggahan, jumlah suka, balasan, hingga informasi pengguna seperti *username* serta lokasi jika tersedia. Hasil dari proses *scraping data* yang telah dikumpulkan tersebut dapat dilihat pada Gambar 4.2 dan Gambar 4.3 berikut

convers	created_at	f	full_text	id_str	image_url	in_reply_to_	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
1,93E+18	Fri May 30 18:3	29	@_upilicious @tanyakanrl A	1,929E+18		_upilicious	in	Mojoagung, Indo	0	2	0	https://x.com/undefinec	8,758E+17	buahhapaini
1,93E+18	Fri May 30 18:0	0	ih anjirlah gw takut ama yg n	1,929E+18			in	Ambon	0	0	0	https://x.com/undefinec	1,646E+18	lifeissemrawut
1,93E+18	Fri May 30 18:0	0	So that's why menurutku bu	1,929E+18		absltlyy	in		0	1	0	https://x.com/undefinec	9,21E+17	absltlyy
1,93E+18	Fri May 30 18:0	2	@meinmokhtar Hang nikah	1,929E+18		meinmokhtar	in	Baling-Kulim-La	0	0	0	https://x.com/undefinec	7,638E+17	rayzarayhan77
1,93E+18	Fri May 30 17:5	0	That's why kalau sebelum ni	1,929E+18		absltlyy	in		0	1	0	https://x.com/undefinec	9,21E+17	absltlyy
1,93E+18	Fri May 30 17:5	0	maaf ya mungkin di bayang	1,929E+18			in	Pekanbaru	0	1	0	https://x.com/undefinec	1,02E+18	ansyrn_
1,93E+18	Fri May 30 17:4	0	Capeeee bgt Tuhan hari ha	1,929E+18			in	nan	0	0	0	https://x.com/undefinec	7,396E+17	nindy_otherside
1,93E+18	Fri May 30 17:3	7	15. bbrp hari stelah MS nikah	1,929E+18		tutorxyim	in	she/her	0	1	0	https://x.com/undefinec	1,256E+18	tutorxyim
1,93E+18	Fri May 30 16:5	0	takut pasti ada tapi kl sama	1,928E+18			in	Jayapura	0	0	0	https://x.com/undefinec	1,892E+18	moriyamom1ji
1,93E+18	Fri May 30 16:5	3	@tanyakanrl aku jg team dil	1,928E+18		tanyakanrl	in	Bandung	0	1	0	https://x.com/undefinec	1,283E+18	emaaoct
1,93E+18	Fri May 30 16:3	0	@tanyarlfes lyaa ya ampun	1,928E+18		tanyarlfes	in	Denpasar	0	0	0	https://x.com/undefinec	1,593E+18	paacarbright
1,93E+18	Fri May 30 16:11	1	takut reflek nyebar undanga	1,928E+18			in	Bandung	0	0	1	https://x.com/undefinec	1,458E+18	nellipeg

Gambar 4.2 Hasil *Scraping* Data Indonesia

convers	created_at	favori	full_text	id_str	image_	in_reply_t	lang	location	quo	rep	retw	tweet_url	user_id	username
1,93E+18	Thu May 29 20:2	0	#Quote 150-Sex is the expr	1,9282E+18			en	United States	0	0	0	https://x.com/ur	1,2E+08	SelfishRoman
1,93E+18	Thu May 29 22:2	1	Jolie has no intention of su	1,9282E+18			en	USA	0	0	0	https://x.com/ur	1,4E+07	SheKnows
1,93E+18	Thu May 29 22:0	0	I wasn t just dishonest wit	1,9282E+18	https://pb		en	United States	0	0	0	https://x.com/ur	1,5E+18	RawMotivatic
1,93E+18	Thu May 29 23:5	2	Random atiny tweet: What'	1,9282E+18	https://pb		en	United States	0	0	0	https://x.com/ur	1,9E+07	karaisshort
1,93E+18	Thu May 29 23:0	0	I predict a divorce on the h	1,9282E+18			en	United States	0	0	0	https://x.com/ur	3,2E+09	jacob001_d
1,93E+18	Thu May 29 21:5	266	@dyingscribe The reason h	1,9282E+18		dyingscribe	en	Texas	0	1	2	https://x.com/ur	3,3E+07	mexillentJuni
1,83E+18	Thu Aug 29 23:56	0	Who the fuck at @espn thc	1,8293E+18			en	Texas	0	0	0	https://x.com/ur	9E+17	BadHopSingle
1,83E+18	Thu Aug 29 23:42	27	Started to go heavy on X th	1,8293E+18	https://pb		en	United States	0	8	1	https://x.com/ur	1,3E+18	MsReippuert
1,83E+18	Thu Aug 29 23:1C	0	Why did it decide to do hal	1,8293E+18			en	United States	0	0	0	https://x.com/ur	8,5E+17	millionmere
1,83E+18	Thu Aug 29 23:16	7	Of all the disastrous parts	1,8293E+18			en	Texas	0	1	0	https://x.com/ur	1,2E+18	billbuckleysta
1,83E+18	Thu Aug 29 23:3C	2	All of the nations that are	1,8293E+18			en	United States	0	1	0	https://x.com/ur	1,3E+18	RedactedRosa

Gambar 4.3 Hasil *Scraping* Data Amerika Serikat

Dengan hasil pengumpulan data tersebut, maka diperoleh *dataset* yang cukup representatif untuk dianalisis lebih lanjut dalam memahami persepsi publik terhadap isu *marriage is scary* di dua negara dengan konteks sosial dan budaya yang berbeda.

4. 2. 3. Hasil Pengolahan *Data Preparation*

Pada tahapan ini merupakan tahapan untuk menganalisis serta pembersihan data sebelum dilakukan pelatihan. Tujuan dari pembersihan data ini adalah agar data menjadi siap digunakan dan tidak ada *noise* yang dapat menggangu hasil *training*. Proses pada tahapan dini dilakukan secara terpisah antara *dataset* Indonesia dan *dataset* Amerika Serikat.

a) Penyeleksian *Dataset*

Tahapan yang dilakukan selanjutnya yaitu melakukan penyeleksian *dataset* berdasarkan *filter* kata kunci, bahasa, serta lokasi pada tiap *dataset* yakni Indonesia dan Amerika Serikat. Proses dilakukan seleksi terlebih dahulu agar data penelitian sesuai dengan konteks penelitian yang dilakukan. Proses Seleksi data yang dilakukan tertera seperti pada Gambar 4.4 dan Gambar 4.5 berikut.

```
keywords = ['marriage is scary', 'takut nikah', 'trauma nikah', 'takut', 'cerai', 'pernikahan indonesia']

mask = df['full_text'].str.contains('|'.join(keywords), case=False, na=False)
mask_lang = df['lang'] == 'in'

indo_df = df[mask & mask_lang].copy()

indo_df = indo_df.dropna(subset=['location'])
indo_df = indo_df[indo_df['location'].str.lower() != 'nan']

print(f'Jumlah data setelah seleksi: {len(indo_df)}')
indo_df.head(100)
```

Gambar 4.4 Seleksi Data Berdasarkan Kata Kunci

```
def lokasi_indo(loc):
    if pd.isnull(loc):
        return None
    for norm_loc, variants in location_map.items():
        if any(variant in loc for variant in variants):
            return norm_loc
    return 'lainnya'


indo_df['lokasi_normal'] = indo_df['location_cleaned'].apply(lokasi_indo)

df_indo = indo_df[indo_df['lokasi_normal'].isin(location_map.keys()) | (indo_df['lokasi_normal'] == 'indonesia')].copy()

print(f'Jumlah total data pada dataframe setelah normalisasi: {len(df_indo)}')
```

Gambar 4.5 Seleksi Data Berdasarkan Lokasi

Gambar 4.4 dan Gambar 4.5 tersebut merupakan proses seleksi data berdasarkan kata kunci, bahasa, serta lokasi pada data penelitian yang telah terkumpul. Data yang sesuai dengan *filter* tersebut kemudian disimpan ke dalam file data baru untuk dilakukan proses penelitian selanjutnya. Hasil banyaknya data setelah dilakukannya penyeleksian data pada kedua negara di jelaskan pada Tabel 4.3 berikut.

Tabel 4.3 Jumlah *Dataset* Setelah Seleksi

	Indonesia	Amerika Serikat
Jumlah Data Awal	7.073	4.953
Jumlah Data Setelah Seleksi	4.031	3.465

Dari Hasil seleksi data tersebut, maka *dataset* akhir antara kedua negara menjadi 4031 data untuk Indonesia dan 3465 untuk Amerika Serikat. *Dataset* ini kemudian akan dilakukan pemrosesan berikutnya untuk dilakukan persiapan pelatihan.

b) Pembersihan Pada *Dataset*

Tahapan pembersihan data atau *data cleaning* merupakan langkah krusial dalam proses *preprocessing* data sebelum dilakukan analisis atau pelatihan model. Dalam penelitian ini, pembersihan *dataset* dilakukan untuk memastikan bahwa data yang digunakan benar-benar bersih, relevan, dan siap digunakan dalam proses tokenisasi dan pelatihan model. Proses pembersihan data dapat dijelaskan pada Gambar 4.6 berikut.

```
def clean_text(text):
    text = str(text).lower()
    text = re.sub(r'http\S+|www\S+|pic\.twitter\.com\S+', '', text)
    text = re.sub(r'@\w+', '', text)
    text = re.sub(r'#\w+', '', text)
    text = re.sub(r'[^a-zA-Z0-9\s]', '', text)
    text = re.sub(r'\s+', ' ', text).strip()
    text = re.sub(r'\d+', '', text)
    return text

df_indo['clean_text'] = df_indo['full_text'].apply(clean_text)
df_indo.drop_duplicates(subset='clean_text', inplace=True)
df_indo.dropna(subset=['clean_text'], inplace=True)
```

Gambar 4.6 Pembersihan data

Proses pembersihan ini seperti pada Gambar 4.6 bahwa mencakup penghapusan data duplikat, *missing value,* menghilangkan karakter yang tidak relevan seperti emoji, simbol, URL, *tag* pengguna (*@username*), tanda baca, angka, serta *hashtag* yang tidak memberikan nilai informasi terhadap sentimen yang dianalisis. Selain itu, dilakukan pula normalisasi teks seperti mengubah seluruh huruf menjadi huruf kecil (*lowercasing*) agar seragam.

Langkah pembersihan juga mencakup penghapusan entri data yang mengandung teks kosong atau bahasa yang tidak sesuai dengan *filter* yang telah ditentukan pada saat proses pengambilan data, yaitu bahasa Indonesia dan Inggris.

Pembersihan ini penting dilakukan untuk mengurangi *noise* dalam data dan meningkatkan kualitas masukan (input) ke dalam model BERT, sehingga proses pelatihan dapat berlangsung secara optimal dan hasil analisis sentimen menjadi lebih akurat. Hasil dari pembersihan *dataset* tersebut dapat dilihat pada Gambar 4.7 dan Gambar 4.8.

	full_text	clean_text
0	Marriage is scary Gimana kalo aku yang sejak k...	marriage is scary gimana kalo aku yang sejak k...
1	@Sentjoko @allah nabi itu ga mau istrinya ribu...	nabi itu ga mau istrinya ribut makanya bersump...
2	Mi madre cerita pas kmarin bude ke rumah si ba...	mi madre cerita pas kmarin bude ke rumah si ba...
4	Random kalo nikah takut banget punya anak cewe...	random kalo nikah takut banget punya anak cewe...
5	@_upilicious @tanyakanrl Aku pernah nemenin da...	aku pernah nemenin dari melompong sampe bantu...

Gambar 4.7 Hasil Pembersihan *Dataset* Indonesia

	full_text	clean_text
0	#Quote 150-Sex is the expression of #love not ...	sex is the expression of not its cause romance...
1	Jolie has no intention of sugar-coating how to...	jolie has no intention of sugarcoating how tou...
2	I wasn t just dishonest with her I was deceiv...	i wasn t just dishonest with her i was deceivi...
3	I predict a divorce on the horizon	i predict a divorce on the horizon
4	@dyingscribe The reason he didnt leave her was...	the reason he didnt leave her was because he k...

Gambar 4.8 Hasil Pembersihan *Dataset* Amerika Serikat

Setelah melalui proses pembersihan data, jumlah *dataset* pada tiap-tiap negara menjadi berkurang. Hal ini menunjukkan bahwa data telah bersih. Sebaran data setelah melakukan pembersihan jumlahnya menjadi seperti pada Tabel 4.4 berikut.

Tabel 4.4 Jumlah Data Setelah Pembersihan

	Indonesia	Amerika Serikat
Jumlah Data Setelah Seleksi	4.031	3.465
Jumlah Data Setelah Pembersihan	3.988	3.359

c) Penghapusan Kolom yang Tidak diperlukan

Tahapan ini merupakan tahapan untuk menyeleksi kolom yang diperlukan saja dalam proses pelatihan. Kolom yang tidak diperlukan akan di buang agar tidak memengaruhi hasil pelatihan. Proses ini dilakukan dengan cara memilih kolom mana saja yang akan tetap dipertahankan dalam data. Kemudian kolom yang tidak dipakai akan dilakukan penghapusan. Sehingga akhir dari data hanya memiliki kolom data yang sesuai dengan proses penelitian. Proses penghapusan kolom pada

data dijelaskan pada Gambae 4.9 berikut.

```
df_indo = df_indo.drop(columns=['conversation_id_str',
       'id_str', 'image_url', 'in_reply_to_screen_name',
       'tweet_url', 'user_id_str', 'full_text'])

print(f'Kolom setelah dihapus:', df_indo.columns)
```

```
Kolom setelah dihapus: Index(['created_at', 'favorite_count', 'lang', 'location', 'quote_count',
       'reply_count', 'retweet_count', 'username', 'location_cleaned',
       'lokasi_normal', 'clean_text'],
      dtype='object')
```

```
df_indo=df_indo.rename(columns={'created_at': 'date'})
df_indo.head()
```

Gambar 4.9 Pemilihan Kolom yang Relevan

Dari Gambar 4.9 bahwa kolom yang dipilih merupakan kolom yang akan digunakan untuk pemrosesan data selanjutnya. Kolom yang dipilih tersebut seperti *'created_at', 'favorite_count', 'lang', 'location', 'quote_count', 'reply_count', 'retweet_count', 'username', 'location_cleaned', 'lokasi_normal', 'clean_text'*.

d) Pembagian *Dataset* Untuk Pelabelan

Pada tahapan ini, *dataset* dari kedua negara, yaitu Indonesia dan Amerika Serikat yang telah melalui proses pembersihan, kemudian dibagi menjadi dua *subset* utama. Proses pembagian dilakukan secara terpisah untuk masing-masing negara agar menjaga keseimbangan proporsi data serta mempertahankan konteks sentimen yang terkandung dalam setiap data. Pembagian dilakukan dengan teknik *random* sampling, yakni data akan di pisah dengan secara acak pada kedua *subset*. Tahapan ini bertujuan untuk memisahkan data menjadi *dataset* untuk dilakukan proses. Proses Pembagian *dataset* untuk pelabelan tersebut terdapat pada Gambar 4.10 berikut.

```
MIN_DATA = 3980
if len(data) < MIN_DATA:
    raise ValueError(f"Data kurang.")

output_dir = "/content/"
filenames = {
    'subset_2000': os.path.join(output_dir, 'subset_2000_indo.csv'),
    'subset_1988': os.path.join(output_dir, 'subset_1988_indo.csv')
}

data_shuffled = data.sample(frac=1, random_state=42).reset_index(drop=True)

split_point = 2000
data_2000_id = data_shuffled.iloc[:split_point]
data_1988_id = data_shuffled.iloc[split_point:split_point + 1988]

data_2000_id.to_csv(filenames['subset_2000'], index=False)
data_1988_id.to_csv(filenames['subset_1988'], index=False)

print(f"Berhasil menyimpan:\n- {filenames['subset_2000']}\n- {filenames['subset_1988']}")
```

Gambar 4.10 Pembagian *Dataset* untuk Pelabelan

Proses pembagian data ini dilakukan pada Gambar 4.10. Data akan di ukur terlebih dahulu dengan jumlah maksimal, kemudian menentukan jumlah data paling banyak untuk digunakan dalam proses pelabelan, yakni dengan jumlah 2000. Setelah itu data akan diacak terlebih dahulu untuk menghilangkan kemungkinan adanya bias pada urutan. Setelah data diacak, proses selanjutnya data akan dibagi menjadi dua *subset* berdasarkan jumlah baris, yakni sebanyak 2000 data pertama serta 1988 data untuk data berikutnya. Proses pembagian ini dilakukan secara manual berdasarkan indeks baris setelah proses pengacakan. Setelah proses pembagian data tersebut, maka jumlah data pada tiap *subset* menjadi sebanyak berikut pada Tabel 4.5.

Tabel 4.5 Jumlah Data *Subset* Penelitian

	Indonesia	Amerika Serikat
Subset 1	2.000	2.000
Subset 2	1.988	1.359
Total	3.988	3.359

Setelah pembagian menjadi dua *subset* tersebut, maka kedua *subset* dengan jumlah masing-masing data sebanyak 2.000 akan digunakan untuk proses pelatihan. Data *subset* 2.000 akan digunakan untuk proses selanjutnya, yaitu pelabelan pada data.

e) Pelabelan Data

Proses pelabelan dalam penelitian ini dilakukan secara otomatis menggunakan model *pretrained* BERT yang disesuaikan dengan bahasa dari masing-masing *dataset*. Untuk data Indonesia, digunakan model *Aardiiiiy/indobertweet-base-Indonesian-sentiment-analysis* yang dikembangkan khusus untuk analisis sentimen pada media sosial berbahasa Indonesia. Proses pelabelan data Indonesia dijelaskan pada Gambar 4.11 berikut.

```python
def predict_sentiment_with_confidence(text):
    inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True)
    with torch.no_grad():
        outputs = model(**inputs)
        logits = outputs.logits
        probs = F.softmax(logits, dim=1).squeeze()

        pred_label_id = torch.argmax(probs).item()
        confidence = probs[pred_label_id].item()

        if pred_label_id == 0:
            label = "negative"
        else:
            label = "positive"

        return label, confidence
```

Gambar 4.11 Pelabelan Data

Sementara itu, data dari Amerika Serikat dilabeli menggunakan model *finiteautomata/bertweet-base-sentiment-analysis*, yaitu model BERTweet yang telah dilatih untuk menganalisis sentimen berbahasa Inggris, khususnya dari media sosial. Proses pelabelannya dilakukan dengan metode yang sama seperti data Indonesia, yaitu melalui tokenisasi, prediksi label sentimen, dan pencatatan hasil prediksi ke dalam *dataset*. Hasil dari pelabelan tersebut terdapat pada Gambar 4.12 dan Gambar 4.13 sebagi berikut.

text	sentiment	confidence
bagus didebatin sebelum nikah biar ketauan dia...	negative	0.995308
liat yg begini selalu kepengen nikah tp masih ...	negative	0.987464
yg belum nikah byk yg takut ini yg udh cerai p...	negative	0.994931
aku pas nikah langsung kasih baju kebaya rok s...	positive	0.665017
bukan tp mingdep temen kecil gue mau nikah t...	negative	0.995761

Gambar 4.12 Hasil Pelabelan *Dataset* Indonesia

text	sentiment	confidence
a good marriage wont work with this ownership	negative	0.853928
barely made it to years of age marriage isnt ...	negative	0.730417
the pattern noticer the middle east north afri...	negative	0.976309
on their wedding anniversary their fighting gl...	negative	0.604531
marriage is not simply for people between the ...	positive	0.942887

Gambar 4.13 Hasil Pelabelan *Dataset* Amerika Serikat

Gambar 4.12 dan Gambar 4.13 merupakan hasil dari pelabelan pada *dataset* Indonesia dan Amerika Serikat. Hasil pelabelan tersebut terkumpul menjadi dua kelas sentimen dengan sebaran jumlah datanya yakni pada Tabel 4.6 untuk *dataset* Indonesia dan Tabel 4.7 untuk *dataset* Amerika Serikat sebagai berikut.

Tabel 4.6 Jumlah Sentimen *Dataset* Indonesia

Indonesia	
Sentimen	Jumlah Data
Positive	212
Negative	1.788
Total	2.000

Tabel 4.7 Jumlah Sentimen *Dataset* Amerika Serikat

Amerika Serikat	
Sentimen	Jumlah Data
Positive	1.178
Negative	821
Total	2.000

Pada Tabel 4.6 diketahui bahwa dari hasil pelabelan tersebut, *dataset* mengalami ketidakseimbangan kelas dengan data condong kepada sentimen negatif dengan jumlah yang cukup besar. Sehingga untuk proses berikutnya dibutuhkan untuk proses *balancing* data terlebih dahulu sebelum melakukan pelatihan model, agar hasil pelatihan menjadi lebih optimal.

f) Penanganan *Data Imbalanced*

Setelah melakukan pelabelan, diketahui bahwa pada *dataset* Indonesia mengalami ketidakseimbangan data, dengan jumlah yang cukup signifikan. Keadaan ketidakseimbangan kelas label ini tidak bisa langsung digunakan pada pelatihan karena akan menimbulkan model bias dalam mempelajari data. Maka dari itu, untuk mengatasi hal ini, dilakukan augmentasi data. Jenis augmentasi data yang digunakan adala *fill-mask* dengan model *indobenchmark/indobert-base-p1*.

Proses tersebut awali dengan mengisi token ke dalam kalimat secara acak dengan token khusus. Kemudian model memprediksi kata yang paling mungkin mengisi posisi token tersebut berdasarkan konteks sekitarnya. Hasil dari proses tersebut akan menghasilkan jenis kalimat baru hasil augmentasi. Setelah melakukan teknik augmentasi, maka *dataset* Indonesia jumlahnya menjadi seimbang dan siap

untuk digunakan dalam pelatihan model. Jumlah dari hasil augmentasi tersebut dapat dijelaskan pada Tabel 4.8 berikut.

Tabel 4.8 Jumlah Data Indonesia Setelah Augmentasi

Sentimen	Jumlah Data
Positive	1.788
Negative	1.788
Total	3.576

Pada Tabel 4.8 tersebut, terlihat bahwa jumlah data pada *dataset* Indonesia telah seimbang. Maka dari itu, setelah proses penyeimbangan data tersebut, maka data telah siap untuk digunakan dalam pelatihan model.

g) Pembagian Data untuk Pelatihan

Tahapan ini merupakan tahapan untuk membagi *dataset* yang siap digunakan untuk pelatihan menjadi bagian-bagian *subset* untuk pelatihan. Pembagian data ini dibagi menjadi 3 (tiga) subset yakni untuk *train data, validation data,* dan *test data.* Pembagian data tersebut dilakukan dengan menggunakan *stratify.* Penggunaan tersebut bertujuan agar pembagian data dilakukan secara sama tiap label sentimennya. Proses pembagian data dijelaskan pada Gambar 4.14 berikut.

```
trainval_indo, test_indo = train_test_split(indo_balanced, test_size=0.1, random_state=42, stratify=indo_balanced['sentiment'])
train_indo, val_indo = train_test_split(trainval_indo, test_size=0.1111, random_state=42, stratify=trainval_indo['sentiment'])

trainval_usa, test_usa = train_test_split(df_usa, test_size=0.1, random_state=42, stratify=df_usa['sentiment'])
train_usa, val_usa = train_test_split(trainval_usa, test_size=0.1111, random_state=42, stratify=trainval_usa['sentiment'])
```

Gambar 4.14 Pembagian *Dataset* untuk Pelatihan

Pembagian *dataset* menjadi tiga jenis data yakni *train data, validatin data,* dan *testing data,* proses ini dilakukan dengan menerapkan teknik *stratify* dengan memanfaatkan fungsi *train_test_split* dari pustaka *scikit-learn.* Hasil dari pembagian data tersebut untuk data indonesia dijelaskan pada Tabel 4.9 berikut.

Tabel 4.9 Jumlah Data *Train*, *Val, Test* di *Dataset* Indonesia

Indonesia			
Sentiment	*Training Data*	*Validation Data*	*Testing Data*
Positif	1.430	179	179
Negatif	1.430	179	179
Total	2.860	356	356

Selain dari *dataset* Indonesia, dataset Amerika Serikat juga dilakukan pembagian data dengan jumlah *subset* yang sama dan dengan teknik yang sama. Hasil pembagian data Amerika Serikat tersebut dijelaskn pada Tabel 4.10 berikut.

Tabel 4.10 Jumlah Data *Train*, *Val, Test* di *Dataset* Amerika Serikat

Amerika Serikat			
Sentiment	*Training Data*	*Validation Data*	*Testing Data*
Positif	943	118	118
Negatif	657	82	82
Total	2.860	300	300

Hasil Dari pembagian data tersebut menujukkan proporsi baik dengan teknik *stratify.* Sehingga hasil dari pembagian data ini menggunakan rasio perbandingan yakni 80:10:10, yakni 80% untuk data *train,* 10% untuk data *validation* dan data *testing.*

h) Hasil Tokenisasi Data

Proses Tokenisasi ini menggunakan tokenisasi berasal dari model yang akan digunakan dalam pelatihan, yakni *pretrained multilangual* BERT. Proses tokenisasi dilakukan dengan memanfaatkan *AutoTokenizer* dari *Transformers.* Hasil Dari Tokenisasi ini adalah berupak kumpulan data yang dipahami oleh model, dimana merupakan *transform* dari data-data teks pada *dataset*.

Proses transformasi data tersebut dilakukan karena agar model tetap dapat memahami bagian-bagian dari teks data yang belum pernah dilihat oleh model sebelumnya, selama bagian-bagian lainnya masih dikenali oleh model. Hasil dari tokenisasi ini misalnya dapat dilihat pada Gambar 4.15 berikut.

```
tokenized = tokenizer(original_text, max_length=16, truncation=True, padding="max_length")
print("Input IDs:", tokenized['input_ids'])
print("Attention Mask:", tokenized['attention_mask'])
```

```
Input IDs: [101, 43024, 17736, 10414, 28977, 39429, 19986, 15371, 10679, 10120, 91196, 11924, 30518, 19503, 191, 102]
Attention Mask: [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```

Gambar 4.15 Hasil Auto Tokenisasi Data

Pada Gambar 4.15 tersebut merupakan hasil dari salah satu proses tokenisasi yakni pada tokenisasi *attention mask* dan *Input IDs.* Proses ini dilakukan pada seluruh data agar model dapat mengenali tiap-tiap bagian teks data, sehingga model dapat dengan mudah untuk melakukan pelatihan.

i) Hasil Transformasi Data Sebelum Pelatihan

Tahapan ini merupakan tahapan lanjutan setelah tokenisasi yakni mengubah data agar dapat digunakan. Tahapan yang dilakukan adalah dengan memastikan format dan strukturnya sesuai dengan kebutuhan model. Transformasi ini bertujuan agar nantinya proses pelatihan model dapat berlangsung secara baik.

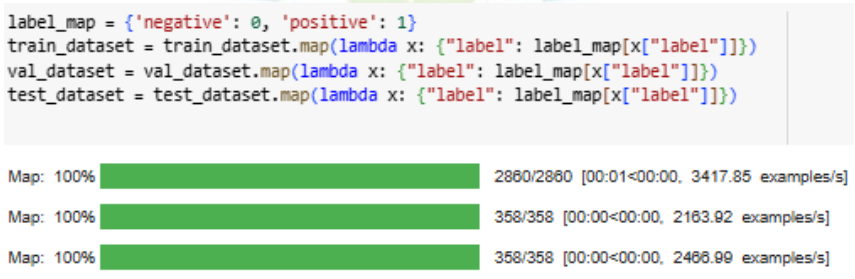
1) Hasil pengubahan kolom *sentiment* menjadi *label*

Transformasi data yang dilakukan pada tahap ini adalah mengubah nama kolom *sentiment* pada *dataset* menjadi nama *label.* Hal ini dilakukan karena penyesuaian sebelum pelatihan karena mengacu pada konvensi umum dari *library transformers* dan *datasets* dari *hugging face.* Hal ini karena kolom target klasifikasi harus dinamakan *label* agar dikenali secara otomatis oleh *pipeline* pelatihan model.

Selain itu, dilakukan penghapusan kolom *text* dari masing-masing *subset* data, yakni *train data, val data,* dan *test data.* Hasil dari penghapusan ini dilakukan pada seluruh teks setelah ditokenisasi menjadi bentuk numerik, karena kolom mentah sudah tidak diperlukan lagi.

2. Hasil konversi *label* menjadi bentuk numerik

Setelah kolom target diubah namanya menjadi *label,* hasil dari tahap berikutnya adalah mengonversi nilai *label* yang awalnya dalam bentuk *string* menjadi bentuk numerik. Hasil dari konversi ini *label positive* menjadi nilai 1 dan *label negative* menjadi nilai 0. Proses pengubahan data *label* menjadi numerik dijelaskan pada Gambar 4.16 berikut.



Gambar 4.16 Konversi Label Menjadi Numerik

Proses penggantian data label pada Gambar 4.16 dilakukan dengan membuat *mapping* label kemudian *mapping* label tersebut diterapkan kepada seluruh data pada data *train, validation,* dan *test.* Hal ini dilakukan karena sebagai bentuk transformasi data untuk menyeseuaikan ketentuan dari pustaka *hugging face datasets.*

4. 2. 4. *Modeling*

Proses *modeling* dilakukan dengan beberapa tahapan untuk melakukan proses pelatihan model untuk pengklasifikasian sentimen. Tahapan dalam pelatihan dilakukan dengan optimal agar mendapatkan hasil model yang baik dalam mengklasifikasikan sentimen.

a) Instalasi Model Pelatihan

Sebelum melakukan *fine-tuning* pada pelatihan. Proses yang dilakukan adalah menginstalasi model pelatihan yang akan digunakan yakni *bert-base-multilangual-cased.* Hasil dari instalasi model pelatihan *bert-base-multilangual-cased* dijelaskan pada Gambar 4.17 berikut.

Gambar 4.17 Hasil Instalasi Model

Pelatihan data dilakukan dengan menggunakan *pretrained* BERT yang telah dapat mengenali 104 bahasa berbeda. Gambar 4.17 model telah berhasil di muat, sehingga tahapan selanjutnya adalah melakukan *fine-tuning* pada pelatihan.

b) Hasil *Fine-tuning* Pelatihan Data

Proses *fine-tuning* dilakukan pada bagian *TrainerArguments* dengan menerapkan beberapa ketentuan pada tiap pelatihan antara dataset Indonesia dan Amerika serikat. Pelatihan model pada *dataset* dilakukan secara terpisah dengan masing-masing data dilakukan beberapa percobaan.

1) *Fine-tuning* Data Indonesia

Jenis-jenis parameter yang digunakan dalam pelatihan pada data indonesia menggunakan beberapa *hyperparameter*, namun yang membedakannya hanya pada jumlah *epoch* pada setiap pelatihan. Pelatihan pada data indonesia dilakukan dengan 3 jenis eksperimen berbeda berdasarkan jumlah *epoch*-nya. Hasil dari pelatihan pada tiap-tiap eksperimen tersebut menunjukkan hasil yang berbeda tiap pelatihan.

a. Eksperimen dengan 3 *epoch*

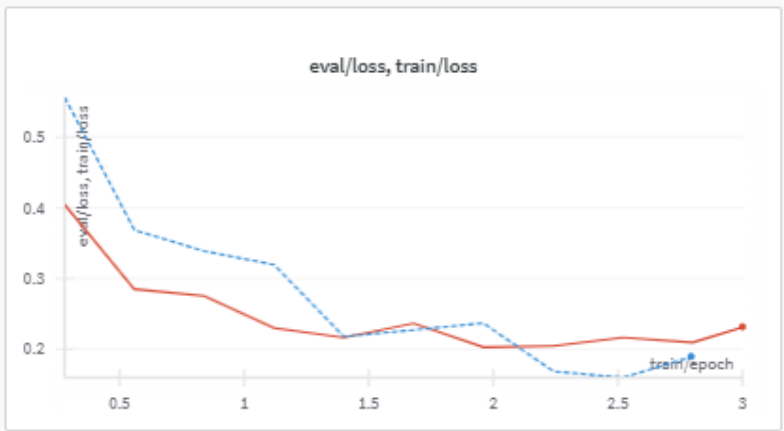
Hasil dari percobaan dengan menggunakan 3 *epoch* dilakukan dengan sebanyak 537 *steps.* Hasil dari performa pelatihan model menunjukkan hasil yang baik, dimana model menunjukkan penurunan *training loss* dan *validation loss* secara konsisten seiring bertambahnya jumlah *step.* Hasil dari pelatihan pada percobaan 3 *epoch* ini dijelaskan pada Gambar 4.18 berikut.

[537/537 12:10, Epoch 3/3]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.556900	0.404851	0.826816	0.831828	0.826816	0.826159
100	0.368600	0.284885	0.891061	0.895518	0.891061	0.890754
150	0.338800	0.275360	0.899441	0.909460	0.899441	0.898822
200	0.319600	0.229797	0.921788	0.927120	0.921788	0.921543
250	0.217800	0.216270	0.927374	0.929304	0.927374	0.927293
300	0.227100	0.236306	0.913408	0.913421	0.913408	0.913407
350	0.236800	0.202625	0.927374	0.933929	0.927374	0.927099
400	0.168600	0.204280	0.938547	0.938767	0.938547	0.938540
450	0.159500	0.216314	0.930168	0.932449	0.930168	0.930075
500	0.189500	0.209479	0.932961	0.933177	0.932961	0.932953

Gambar 4.18 Hasil *Training* Eksperimen Pertama Indonesia

Dalam Gambar 4.18, pada langkah ke-500*, training loss* mencapai 0.1895 dan *validation loss* sebesar 0.2049. Selain itu grafik *loss pada* Gambar 4.19 yang menandakan bahwa model telah belajar dengan baik dan tidak mengalami o*verfitting.* Hasil grafik *loss* evaluasi pada percibaan pertama pada data Indonesia dapat dijelaskan pada Gambar 4.19 berikut.



Gambar 4.19 Grafik *Loss* Eksperimen Pertama Indonesia

b. Eksperimen dengan 5 *epoch*

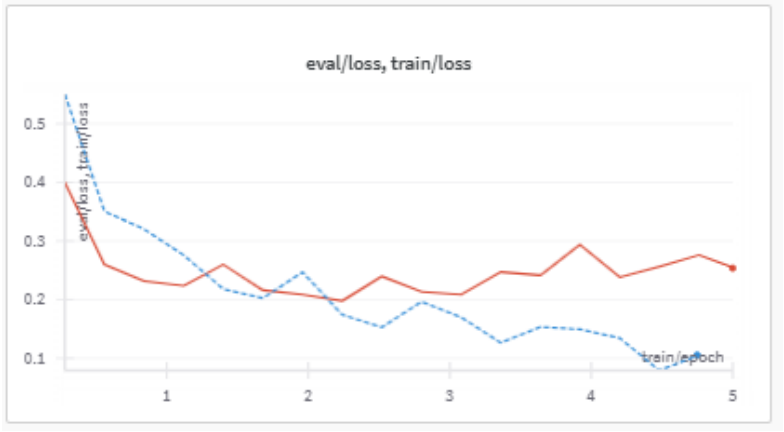
Eksperimen pada data Indonesia dilakukan dengan menggunakan 5 *epoch,* dengan pelatihan dilakukan menggunakan 850 *steps.* Hasil dari pelatihan ini menunjukkan hasil yang baik, dimana adanya penurunan *training loss* yang konsisten dari awal hingga akhir *epoch.* Kemudian nilai akhir mencapai 0.1054 pada *step* ke-850. Hasil pelatihan pada percobaan dengan 5 *epoch* pada data Indonesia dijelaskan pada Gambar 4.20 berikut.

[895/895 18:38, Epoch 5/5]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.550700	0.399752	0.818436	0.824771	0.818436	0.817546
100	0.350600	0.260342	0.896648	0.899453	0.896648	0.896466
150	0.320100	0.231708	0.916201	0.918080	0.916201	0.916107
200	0.275800	0.223768	0.921788	0.924384	0.921788	0.921668
250	0.218100	0.260005	0.916201	0.916669	0.916201	0.916178
300	0.202500	0.215956	0.927374	0.928230	0.927374	0.927338
350	0.247000	0.208450	0.921788	0.927120	0.921788	0.921543
400	0.174200	0.198096	0.935754	0.936422	0.935754	0.935730
450	0.152800	0.239663	0.935754	0.936094	0.935754	0.935742
500	0.196500	0.213511	0.941341	0.942459	0.941341	0.941304
550	0.169800	0.208557	0.944134	0.945023	0.944134	0.944106
600	0.126500	0.247264	0.944134	0.944634	0.944134	0.944118
650	0.153400	0.241489	0.938547	0.939425	0.938547	0.938517
700	0.149400	0.293748	0.921788	0.922632	0.921788	0.921749
750	0.134500	0.237993	0.938547	0.939425	0.938547	0.938517
800	0.079200	0.256112	0.938547	0.939425	0.938547	0.938517
850	0.105400	0.276145	0.946927	0.946941	0.946927	0.946927

Gambar 4.20 Hasil *Training* Data Eksperimen Kedua Indonesia

Selain itu hasil *validation loss* menunjukkan hasil yang cenderung fluktuatif akan tetapi stabil dalam kisaran rendah, hasil nilai akhirnya sebesar 0.2761. Hasil tersebut dapat dilihat pada Gambar 4.20 dan grafik loss pada Gambar 4.21 berikut.



Gambar 4.21 Grafik *Loss* Percobaan Kedua Indonesia

c. Eskperimen dengan 8 *epoch*

Hasil dari percobaan pelatihan ketiga dilakukan dengan menggunakan 8 *epoch.* Dari hasil pelatihan tersebut diketahui bahwa, hasil pelatihan menunjukkan hasil yang baik dengan *training loss* menurun secara konsisten hingga mencapai 0.0630

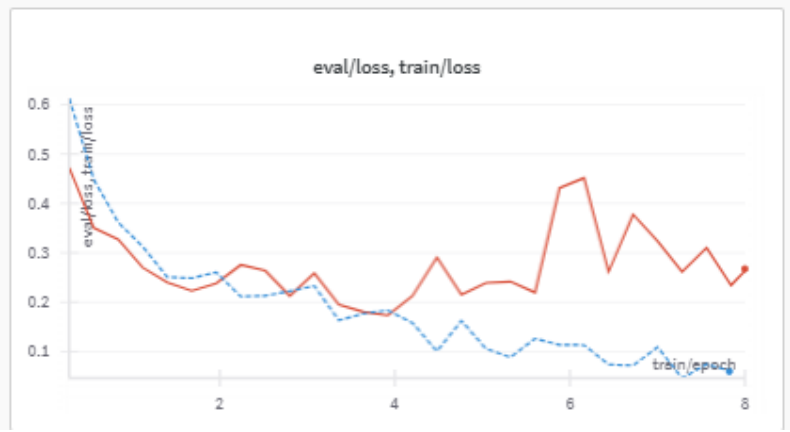
pada *step* ke-1400. Akan tetapi pada *validation loss* mengalami fluktuasi yang cukup signinifikan dibeberapa titik. Hasil pelatihan pada percobaan dengan 8 *epoch* pada data Indonesia dijelaskan pada Gambar 4.22 berikut.

[1432/1432 37:46, Epoch 8/8]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.613200	0.472586	0.815642	0.817067	0.815642	0.815435
100	0.448900	0.350675	0.854749	0.856350	0.854749	0.854585
150	0.361600	0.327336	0.868715	0.881688	0.868715	0.867590
200	0.312600	0.270096	0.899441	0.904491	0.899441	0.899127
250	0.251000	0.239781	0.918994	0.923769	0.918994	0.918766
300	0.248800	0.222973	0.916201	0.920453	0.916201	0.915989
350	0.260700	0.238210	0.921788	0.927120	0.921788	0.921543
400	0.211600	0.275657	0.910615	0.910666	0.910615	0.910612
450	0.213200	0.264073	0.924581	0.926190	0.924581	0.924510
500	0.222700	0.212543	0.935754	0.938836	0.935754	0.935641
550	0.233100	0.258624	0.924581	0.924700	0.924581	0.924576
600	0.163500	0.194839	0.944134	0.945023	0.944134	0.944106
650	0.177000	0.180408	0.949721	0.952489	0.949721	0.949644
700	0.183700	0.173263	0.944134	0.946139	0.944134	0.944071
750	0.158400	0.212673	0.944134	0.945525	0.944134	0.944090
800	0.100800	0.290675	0.932961	0.933015	0.932961	0.932959
850	0.162600	0.214913	0.952514	0.954229	0.952514	0.952469
900	0.106100	0.239348	0.938547	0.938602	0.938547	0.938546
950	0.088600	0.241712	0.946927	0.947612	0.946927	0.946907
1000	0.126600	0.219654	0.949721	0.950621	0.949721	0.949696
1050	0.113600	0.430883	0.905028	0.908290	0.905028	0.904838
1100	0.113500	0.451524	0.910615	0.914809	0.910615	0.910388
1150	0.074000	0.261718	0.941341	0.941355	0.941341	0.941340
1200	0.072000	0.377760	0.918994	0.921216	0.918994	0.918887
1250	0.109300	0.323463	0.924581	0.926190	0.924581	0.924510
1300	0.046500	0.261905	0.938547	0.938767	0.938547	0.938540
1350	0.074600	0.309943	0.930168	0.930826	0.930168	0.930141
1400	0.060300	0.234468	0.952514	0.952641	0.952514	0.952511

Gambar 4.22 Hasil *Training* Data Eksperimen Ketiga Indonesia

Selain itu, pada Gambar 4.23 grafik *loss* dimana *eval loss* sempat meningkat meskipun *train loss* turun. Sehingga hal ini menunjukkan adanya indikasi awal *overfitting* dimana model terlalu menyesuaikan diri terhadap data pelatihan, akan tetapi tidak memberikan peningkatan pada data validasi. Grafik *loss* pada percobaan ketiga pada pelatihan data Indonesia dijelaskan pada Gambar 4.23 berikut.



Gambar 4.23 Grafik *Loss* Percobaan Ketiga Indonesia

2) *Fine-Tuning* Data Amerika Serikat

Pelatihan pada *dataset* Amerika Serikat dilakukan dengan empat eksperimen yang berbeda untuk menguji *fine-tuning* dalam berbagai kondisi. Pada setiap percobaan yang dilakukan memiliki *fine-tuning* yang berbeda. Hasil pada eksperimen tersebut dijelaskan berikut.

a. Eksperimen pertama *fine-tuning* pertama

Hasil dari percobaan pelatihan pada data Amerika Serikat dilakukan dengan menggunakan *fine-tuning* standar. Hasil dari pelatihan menunjukkan hasil yang baik. Pada Gambar 4.24, nilai *accuracy* sebesar 0.77, *precision* sebesar 0.765, *recall* dengan nilai 0.7734, serta nilai *F1-score* sebesar 0.7666. Hasil pelatihan pada percobaan pertama pada data Amerika Serikat dijelaskan pada Gambar 4.24 berikut.

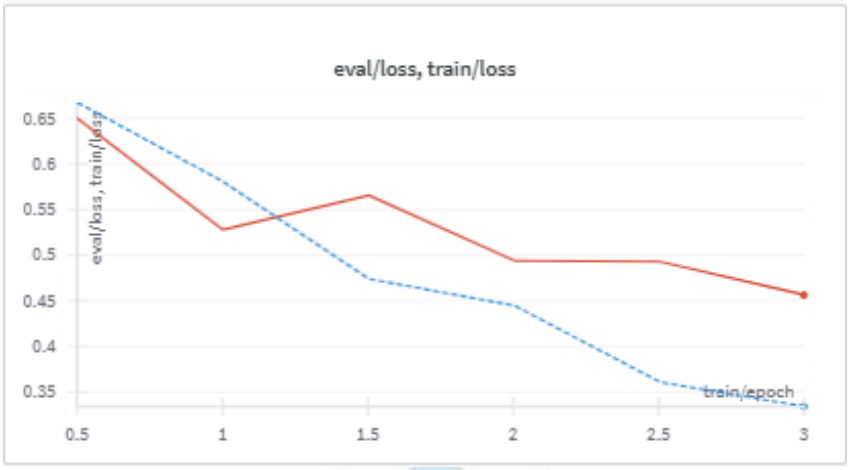
[300/300 06:14, Epoch 3/3]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.667600	0.650049	0.590000	0.295000	0.500000	0.371069
100	0.580800	0.527968	0.725000	0.723965	0.696259	0.700688
150	0.474000	0.565455	0.730000	0.766978	0.760025	0.729757
200	0.444700	0.494142	0.755000	0.749918	0.736565	0.740679
250	0.360600	0.493021	0.780000	0.776843	0.785655	0.777305
300	0.333900	0.494771	0.770000	0.765556	0.773460	0.766640

Gambar 4.24 Hasil Pelatihan Pertama Data Amerika Serikat

Evaluasi kinerja model juga pada Gambar 4.25 berupa grafik *loss* model yang menunjukkan bahwa nilai *train loss* dan *validation loss* sama-sama menurun secara stabil tanpa adanya lonjakan tajam, sehingga menandakan model tidak mengalami *overfitting* serta mampu melakukan generalisasi dengan baik pada data validasi.

Grafik *loss* pada percobaan pertama pada data Amerika Serikat dijelaskan pada Gambar 4.25 berikut.



Gambar 4.25 Grafik *Loss* Pelatihan Pertama Amerika Serikat

b. Eksperimen kedua *fine-tuning* kedua

Percobaan kedua dilakukan dengan *fine-tuning* sedikit berbeda, yakni dengan menambahkan *WeightedTrainer* dengan bobot kelas otomatis dan *early stopping.* Tujuannya *s*upaya model lebih sensitif terhadap kelas minoritas, yakni kelas negatif. Hasil dari pelatihan ini menunjukkan akurasi sebesar 0.78, *F1-score* sebesar 0.77, kemudian *recall* dengan 0.78. Hasil pelatihan pada percobaan kedua pada data Amerika Serikat dijelaskan pada Gambar 4.26 berikut.

[150/500 03:17 < 07:46, 0.75 it/s, Epoch 1/5]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.106600	1.265606	0.805000	0.826708	0.775217	0.785000
100	0.141200	1.361504	0.760000	0.766359	0.774287	0.759133
150	0.021900	1.471083	0.780000	0.773737	0.780074	0.775602

Gambar 4.26 Hasil Pelatihan Kedua Amerika Serikat

Dilihat pada Gambar 4.27 yang merupakan grafik *loss,* meski *training loss* stabil, akan tetapi *validation loss* meningkat, sehingga menngindikasikan *overfitting.* Sehingga secara keseluruhan model ini menurun dibanding percobaan pertama. Grafik *loss* pada percobaan kedua pada pelatihan data Amerika Serikat dijelaskan pada Gambar 4.27 berikut.

Gambar 4.27 Grafik *Loss* Pelatihan Kedua Amerika Serikat

c. Eksperimen *fine-tuning* ketiga

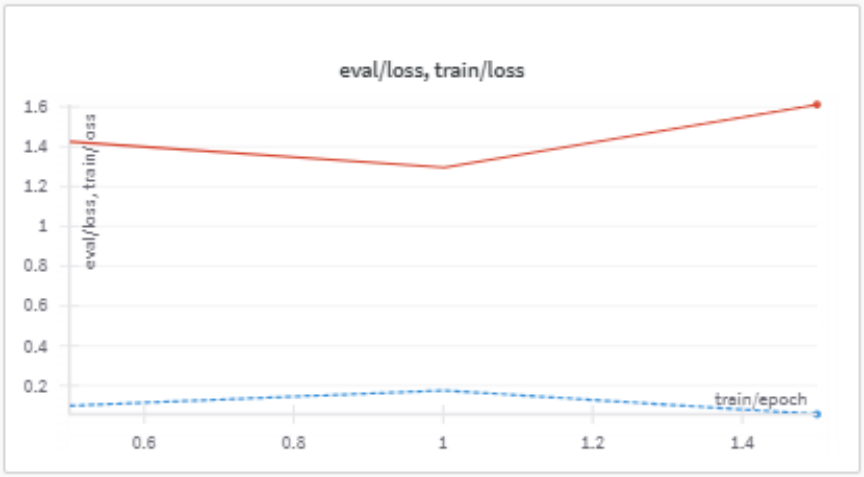
Pelatihan model pada percobaan ketiga memiliki hasil yang relatif stabil serta cukup seimbang antara dua kelas. Proses pelatihan data dilakukan dengan *early stoping* dan pelatihan berhenti di *epoch* 1 karena model tidak memberikan hasil yang signifikan. Hasilnya *validation loss* sebesar 1.575, akurasi sebesar 0.77, *precision* 0.77, *recall* sebesar 0.77, dan F1-score sebesar 0.77. Hasil pelatihan pada percobaan ketiga pada data Amerika Serikat dijelaskan pada Gambar 4.28 berikut.

[150/500 02:58 < 07:02, 0.83 it/s, Epoch 1/5]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.098800	1.424941	0.810000	0.813214	0.810000	0.810875
100	0.175300	1.295842	0.705000	0.781691	0.705000	0.701176
150	0.055500	1.575608	0.770000	0.778888	0.770000	0.771680

Gambar 4.28 Hasil Pelatihan Ketiga Amerika Serikat

Akan tetapi pada Gambar 4.29 merupakan grafik *loss* menunjukkan adanya jarak yang semakin besar antara *training loss* dan *validation loss* setelah epoch ke 1.5, dimana *validation loss* cenderung meningkat meski *training loss* menurun, sehingga menjadi indikasi *overfitting*. Hal ini karena model terlalu menyesuaikan diri dengan data latih dan tidak generalisasi dengan data validasi. Grafik *loss* pada percobaan ketiga pada pelatihan data Amerika Serikat dijelaskan pada Gambar 4.29 berikut.



Gambar 4.29 Grafik Loss Pelatihan Ketiga Amerika Serikat

d. Eksperimen *fine-tuning* keempat

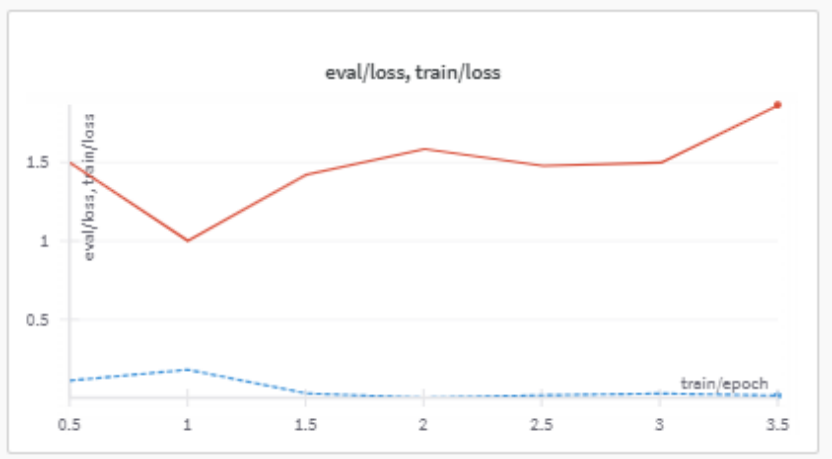
Percobaan keempat dilakukan dengan menggunakan *class weight* manual serta dengan *threshold* prediksi sebesar 0.5. Selain itu, pelatihan menggunakan *early stopping* agar model dapat otomatis berhenti apabila performa tidak menaglami peningkatan. Hasil pelatihan pada percobaan keempat pada data Amerika Serikat dijelaskan pada Gambar 4.30 berikut.

[350/500 07:08 < 03:04, 0.81 it/s, Epoch 3/5]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
50	0.110500	1.499760	0.800000	0.808671	0.800000	0.801461
100	0.180400	1.000648	0.795000	0.807183	0.795000	0.796632
150	0.029600	1.422339	0.805000	0.810675	0.805000	0.806207
200	0.001500	1.584814	0.800000	0.804848	0.800000	0.801142
250	0.017800	1.478273	0.820000	0.820796	0.820000	0.820315
300	0.028900	1.498684	0.815000	0.817495	0.815000	0.815735
350	0.015700	1.566785	0.800000	0.801958	0.800000	0.800657

Gambar 4.30 Hasil Pelatihan Keempat Amerika Serikat

Hasil dari evaluasi menunjukkan bahwa akurasi sebesar 0.80 dan *precision* sebesar 0.80. Serta *recall* sebesar 0.80. Kemudian nilai *F1-score* yang diperoleh adalah 0.80, Dari grafik *loss* pada Gambar 4.31 bahwa terlihat *training loss* menurun, akan tetapi *validation loss* mengalami fluktuasi serta cenderung meningkat. Hal ini menunjukkan indikasi *overfitting.* Meskipun hal tersebut terjadi, model hasil pelatihan ini masih cukup stabil. Grafik *loss* pada percobaan keempat pada pelatihan data Amerika Serikat dijelaskan pada Gambar 4.31 berikut.



Gambar 4.31 Grafik *Loss* Pelatihan Keempat Amerika Serikat

4. 2. Hasil Serta Pembahasan Rumusan Masalah Penelitian 2

Bagian ini merupakan pembahasan mengenai hasil evaluasi model klasifikasi sentimen yang telah dibangun berdasarkan rumusan masalah kedua, yakni berapakah akurasi algoritma BERT dalam analisis sentimen pada data *tweet* mengenai isu "*marriage is scary"*. Evaluasi model hasil pelatihan dilakukan dengan berbagai metrik performa yang mencakup *accuracy, precision, recall,* dan F1-score. Pembahasan juga dilengkapi dengan visualisasi *confusion matrix* untuk melihat distribusi prediksi model.

4. 2. 1. *Evaluation*

Tahapan pada proses ini mengevaluasi secara menyeluruh kualitas model yang telah dibangun. Pada tahap ini model dievaluasi berdasarkan sejumlah metrik performa seperti *accuration, precision, recall,* dan F1-score, serta dengan *confusion matrix.*

a) Hasil Pelatihan Model Indonesia

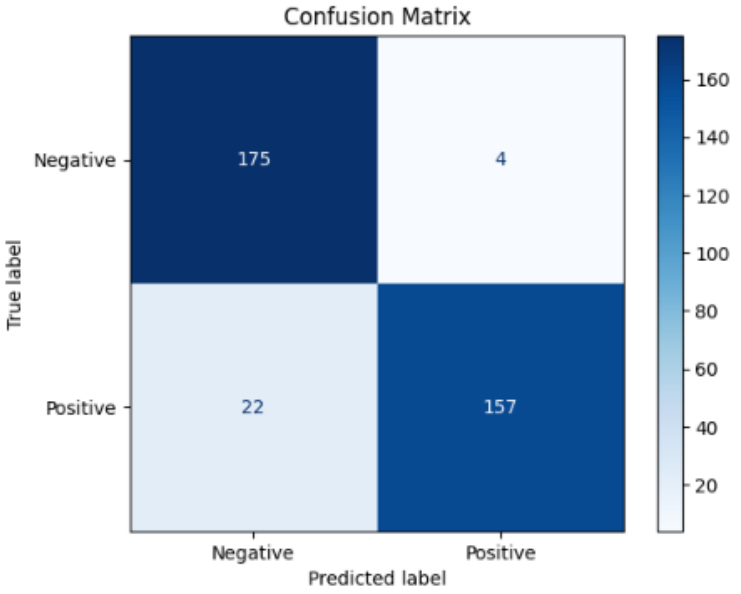
1) Percobaan Pelatihan Pertama dengan 3 *Epoch*

Hasil evaluasi pada percobaan pertama pada data Indonesia ini dilihat pada Tabel 4.11 merupakan *classification report* hasil dari model pada percobaan pertama. Pada pelatihan model pertama ini menunjukkan hasil akurasi sebesar 93%, dengan *precision* 0.93, *recall* sebesar 0.93, dan *F1-score* sebesar 0.93, sehingga hal ini menunjukkan hasil yang sangat baik. Hasil *classification report* pada evaluasi pelatihan data Indonesia pertama dengan menggunakan data uji dijelaskan pada Tabel 4.11 berikut.

Tabel 4.11 *Classification Report* Pelatihan Pertama Indonesia

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.89	0.98	0.93	179
Positive	0.98	0.88	0.92	179
Macro avg	0.93	0.93	0.93	358
Weighted avg	0.93	0.93	0.93	358
accuracy			0.93	358

Selain itu, dilihat dari hasil *confusion matrix* pada Gambar 4.32 menunjukkan hasil prediksi yang cukup baik dimana model berhasil memprediksi kelas negatif sebanyak 175 benar dan 4 salah, dengan *precision* sebesar 0.89 dan *recall* sangat tinggi sebesar 0.98. Kemudian model berhasil memprediksi kelas positif sebanyak 157 benar dan 22 salah, dengan *precison* sebesar 0.98, dan *recall* sebesar 0.88. Dari hasil tersebut bahwa mengindikasikan model sedikit sensitif terhadap kelas negatif, dengan akurasi dan *F1-score* diatas 0.93 untuk kedua kelas. Hasil dari *confusion matrix* pada evaluasi model pelatihan pertama dengan data uji Indonesia dijelaskan pada Gambar 4.32 berikut.



Gambar 4.32 *Confusion Matrix* Pelatihan Pertama Indonesia

Dari hasil tersebut, pelatihan model dengan 3 *epoch* menunjukkan hasil yang baik dengan akurasi total sebesar 93% pada data uji. Sehingga model hasil pelatihan pada percobaan yang pertama memiliki performa yang baik.

2) Percobaan Pelatihan Pertama dengan 5 *Epoch*

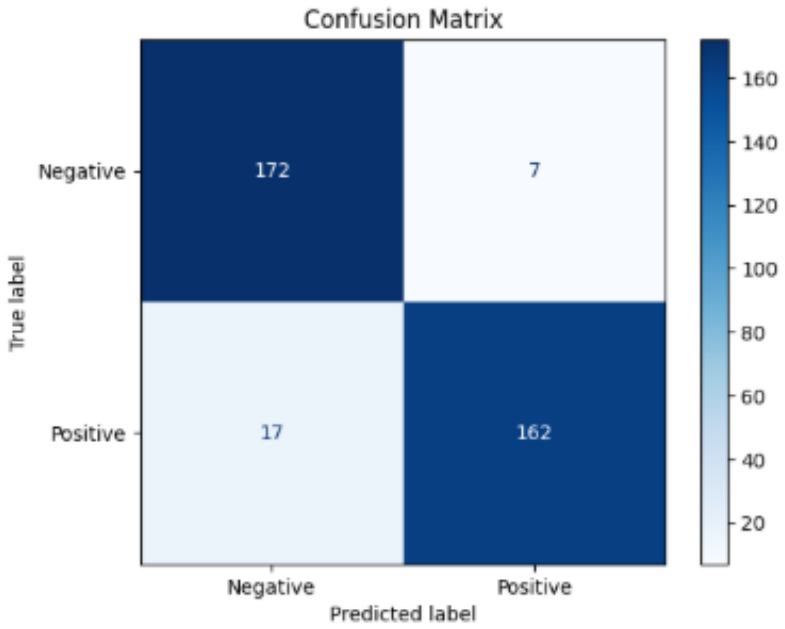
Tabel 4. 12 merupakan *classification report* dari hasil model percobaan kedua.

Hasil dari model menunjukkan hasil yang baik. Hal ini karena hasil akurasi sebesar 93%, *precision* sebesar 0.93, *recall* sebesar *0.93,* dan *F1-score* sebesar 0.93. Hasil *classification report* pada evaluasi pelatihan data Indonesia kedua dengan menggunakan data uji dijelaskan pada Tabel 4.12 berikut.

Tabel 4.12 *Classification Report* Pelatihan Kedua Indonesia

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.91	0.96	0.93	179
Positive	0.96	0.91	0.93	179
Macro avg	0.93	0.93	0.93	358
Weighted avg	0.93	0.93	0.93	358
accuracy			0.93	358

Begitu juga dengan pada Gambar 4.33 yang merupakan *confusion Matrix* dari hasil pelatihan model ini yang menunjukkan hasil prediksi yang baik. Hasilnya dimana model berhasil memprediksi 172 benar dan 7 salah dengan *recall* tinggi sebesar 0.96 dan *precision* baik di nilai 0.91. selain itu untuk prediksi kelas positif model berhasil memprediksi sebanyak 162 benar dan 17 salah, dengan *precision* tinggi 0.96 dan *recall* sedikit lebih rendah sebesar 0.91. Hasil dari *confusion matrix* pada evaluasi model pelatihan kedua dengan data uji Indonesia dijelaskan pada Gambar 4.33 berikut.



Gambar 4.33 *Confusion Matrix* Percobaan Kedua Indonesia

Hasil dari pelatihan model pada percobaan kedua tersebut cenderung lebih rendah dibandingkan dengan hasil pelatihan pada percobaan pertama dalam

memprediksi data uji.

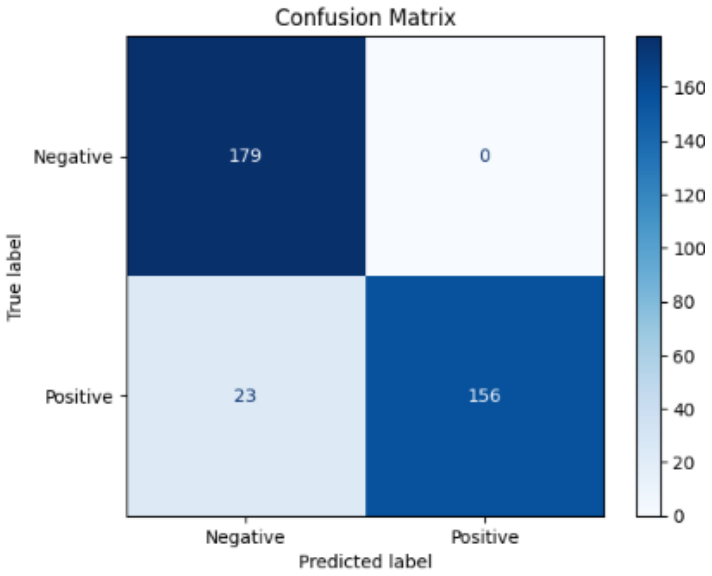
3) Percobaan dengan 8 *epoch*

Hasil evaluasi model pelatihan pada percobaan ketiga, diketahui memiliki hasil yang baik. Pada Tabel 4.13 merupakan hasil prediksi model terhadap data uji terlihat bahwa model berhasil memprediksi data kelas negatif secara sempurna, akan tetapi masih banyak yang salah dalam memprediksi data kelas positif. Kemudian memiliki nilai *accuracy* yang baik sebesar 94% setelah memprediksi data uji. Hasil *classification report* pada evaluasi pelatihan data Indonesia ketiga dengan menggunakan data uji dijelaskan pada Tabel 4.13 berikut.

Tabel 4.13 *Classification Report* Pelatihan Ketiga Indonesia

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.89	1.00	0.94	179
Positive	1.00	0.87	0.93	179
Macro avg	0.94	0.94	0.94	358
Weighted avg	0.94	0.94	0.94	358
accuracy			0.94	358

Selain itu, pada Gambar 4.34 menunjukkan hasil *confusion matrix* dimana model berhasil memprediksi kelas negatif sebanyak 179 dengan 0 salah prediksi. Selain itu model berhasil memprediksi kelas positif sebanyak 156 dengan salah prediksi sebanyak 23. Sehingga model cenderung lebih berhati hati dalam mengklasifikasikan kelas positif, dimana menjadikan banyak data positif yang salah terprediksi sebagai negatif. Hasil dari *confusion matrix* pada evaluasi model pelatihan kedua dengan data uji Indonesia dijelaskan pada Gambar 4.34 berikut.



Gambar 4.34 *Confusion Matrix* Pelatihan Ketiga Indonesia

Dari hasil evaluasi pada pelatihan model dengan data Indonesia tersebut setiap percobaan memiliki kelebihan dan kekurangan masing-masing. Sehingga hasil pelatihan tersebut dapat di jelaskan pada Tabel 4.14 berikut.

Tabel 4.14 Hasil Eksperimen Data Indonesia

Epoch	Kelas	*Precision*	*Recall*	*F1-score*	*Accuracy*	*Over-fitting*	Catatan
3	*Negative*	0.89	0.98	0.93	0.93	Tidak	*Train* & *valid loss* stabil
	Positive	0.98	0.88	0.92			
5	*Negative*	0.91	0.96	0.93	0.93	Tidak	Kinerja paling seimbang
	Positive	0.96	0.91	0.93			
8	*Negative*	0.89	1.00	0.94	0.94	Ada Indikasi	*Valid loss* naik diakhir
	Positive	1.00	0.87	0.93			

Berdasarkan dari ketiga hasil pelatihan tersebut, bahwa diketahui model pada pelatihan pertama yakni dengan 5 *epoch* merupakan model yang lebih baik. Hal ini karena model lebih seimbang dan optimal dalam hal performa dan generalisasi. Sehingga model ini mampu mempertahankan nilai metrik yang tinggi tanpa adanya *overfitting*.

b) Hasil Pelatihan Model Amerika Serikat

1) Percobaan Pertama Pelatihan

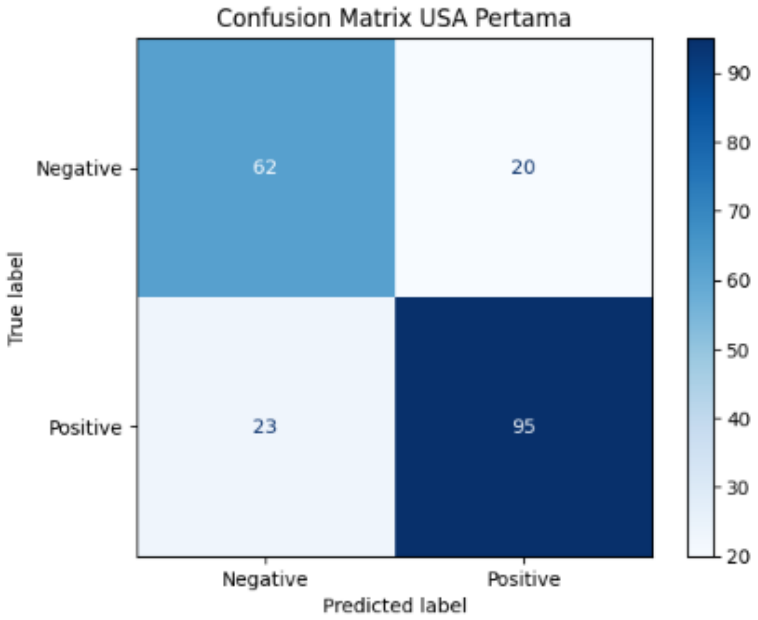
Hasil dari performa model pada percobaan pertama ini menunjukkan hasil yang

cukup baik dengan akurasi 0.79 dan rata-rata F1-score sebesar 0.79 pada Tabel 4.15. Hal ini mengindikasikan model seimbang dalam mengklasifikasikan sentimen negatif maupun positif, meski performa pada kelas negatif sedikit lebih rendah dibandingkan kelas positif. Hasil *classification report* pada evaluasi pelatihan data Amerika Serikat pertama dengan menggunakan data uji dijelaskan pada Tabel 4.15 berikut.

Tabel 4.15 *Classification Report* Pelatihan Pertama Amerika Serikat

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.73	0.76	0.74	82
Positive	0.83	0.81	0.82	118
Macro avg	0.78	0.78	0.79	200
Weighted avg	0.79	0.79	0.78	200
accuracy			0.79	200

Pada Gambar 4.35 yakni merupakan *confusion matrix* dari hasil pelatihan model ini, bahwa model dapat mengklasifikasikan sentimen positif dengan cukup seimbang dengan 62 prediksi benar untuk kelas negatif dan 95 prediksi benar untuk kelas positif. Selain itu jumlah kesalahan klasifikasi yang relatif kecil sebesar 20 *false positive* dan *23 false negative.* Hasil dari *confusion matrix* pada evaluasi model pelatihan pertama dengan data uji Amerika Serikat dijelaskan pada Gambar 4.35 berikut.



Gambar 4.35 *Confusion Matrix* Pelatihan Pertama Amerika Serikat

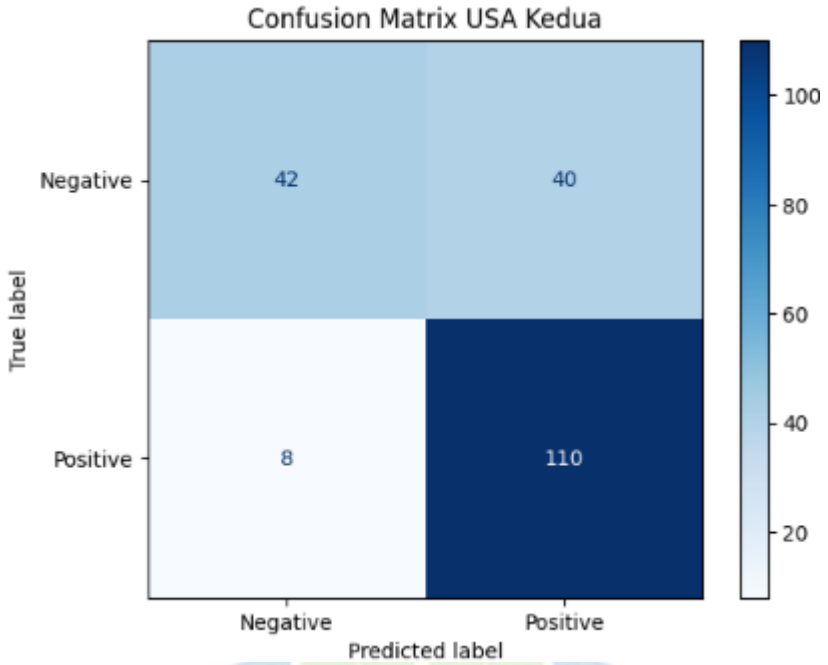
2) Percobaan Kedua Pelatihan

Hasil evaluasi pada model dari percobaan kedua, terlihat bahwa model memiliki performa yang baik. Seperti pada Tabel 4.16 merupakan hasil dari *classification report* pada percobaan kedua ini. Hasil menunjukkan bahwa terdapat *recall* yang kecil pada kelas negatif sebesar 0.51 dibandingkan dengan kelas positif yang lebih besar sebesar 0.93. Kemudian nilai *F1-score* yang kecil pada kelas negatif dengan nilai 0.64 sehingga model sedikit kurang memprediksi kelas negatif dibandingkan dengan kelas positif. Hasil *classification report* pada evaluasi pelatihan data Amerika Serikat kedua dengan menggunakan data uji dijelaskan pada Tabel 4.16 berikut.

Tabel 4.16 *Classification Report* Pelatihan Kedua Amerika Serikat

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.87	0.51	0.64	82
Positive	0.73	0.93	0.82	118
Macro avg	0.79	0.72	0.76	200
Weighted avg	0.78	0.76	0.73	200
accuracy			0.75	200

Maka dari itu, performa model ini cenderung kurang seimbang pada antar kelas. Seperti pada Gambar 4.36 yakni *confusion matrix* menunjukkan nilai yang tidak seimbang. Di mana model sangat baik dalam mengklasifikasin sentimen positif tapi sangan lemah dalam sentimen negatif. Hal ini mengindikasikan adanya kecenderungan terhadap salah satu kelas saja yaitu selalu memprediksi sebagai sentimen positif. Hasil dari *confusion matrix* pada evaluasi model pelatihan kedua dengan data uji Amerika Serikat dijelaskan pada Gambar 4.36 berikut.



Gambar 4.36 *Confusion Matrix* Pelatihan Kedua Amerika Serikat

3) Percobaan Ketiga Pelatihan

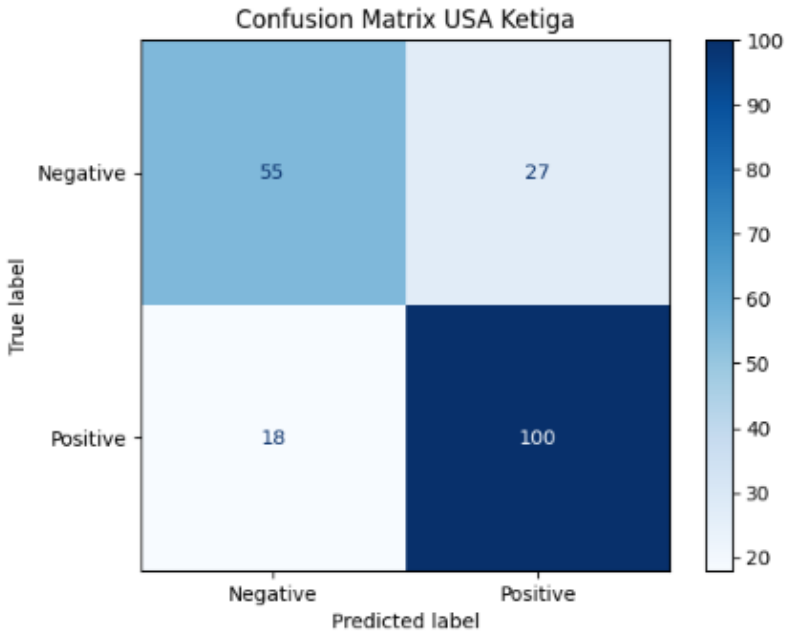
Hasil evalusi model pada pelatihan ketiga terlihat pada Tabel 4.17, dimana menunjukkan hasil bahwa model mencapai akurasi sebesar 78%, dengan *F*1-score macro sebesar 0.78. Pada model ini, model memiliki performa klasifkasi terhadap sentimen positif yang lebih baik dibanding dengan kelas negatif dengan nilai *recall* kelas positif sebanyak 0.85 serta pada kelas negatif sebanyan 0.67. Sehingga model ini masih kurang dalam memprediksi kelas negatif secara baik. Hasil *classification report* pada evaluasi pelatihan data Amerika Serikat ketiga dengan menggunakan data uji dijelaskan pada Tabel 4.17 berikut.

Tabel 4.17 *Classification Report* Pelatihan Ketiga Amerika Serikat

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.75	0.67	0.71	82
Positive	0.79	0.85	0.82	118
Macro avg	0.77	0.76	0.78	200
Weighted avg	0.77	0.78	0.76	200
accuracy			0.78	200

Pada Gambar 4.37 yang merupakan *confusion matrix* dapat dilihat bahwa model berhasil mengklasifikasikan 55 dari 82 data negatif secara benar. Selain itu 100 dari 118 data positif terprediksi secara benar. Selain itu model melakukan kesalahan klasifikasi terhdap 27 data negatif yang terprediksi sebagai positif dan 18

data positif yang diprediksi sebagai negatif. Hasil dari *confusion matrix* pada evaluasi model pelatihan ketiga dengan data uji Amerika Serikat dijelaskan pada Gambar 4.37 berikut.



Gambar 4.37 *Confusion Matrix* Percobaan Ketiga Amrika Serikat

4) Percobaan Keempat Pelatihan

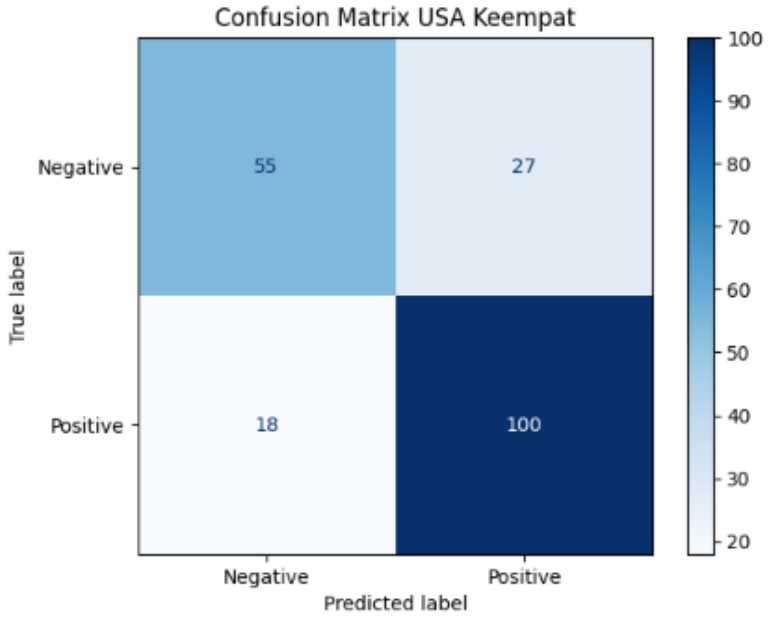
Hasil evaluasi pada percobaan keempat terlihat pada Tabel 4.18 yang menunjukkan model memiliki akurasi 78%, serta nilai *F1*-score makro sebesar 0.78. Sehingga disimpulkan model menunjukkan kinerja yang lebih bagus dalam mengklasifikasikn sentimen positif dibandingkan dengan kelas negatif. Selain itu, nilai model memiliki nilai *precision* sebesar 0.79, *recall* sebesar 0.85, dan *F1-score* sebesar 0.82. Hasil *classification report* pada evaluasi pelatihan data Amerika Serikat keempat dengan menggunakan data uji dijelaskan pada Tabel 4.18 berikut.

Tabel 4.18 *Classification Report* Pelatihan Keempat Amerika Serikat

	Precision	*Recall*	*F1-score*	*Support*
Negative	0.75	0.67	0.71	82
Positive	0.79	0.85	0.82	118
Macro avg	0.77	0.76	0.78	200
Weighted avg	0.77	0.78	0.77	200
accuracy			0.78	200

Kemudian pada Gambar 4.38 merupakan *confusion matrix* menunjukkan bahwa model dapat mengklasifikasi 55 data negatif dan 100 data positif dengan

benar. Akan tetapi masih terdapat 27 data negatif yang salah diklasifikasikan, serta 18 data positif yang salah diklasifikasikan juga. Hasil dari *confusion matrix* pada evaluasi model pelatihan keempat dengan data uji Amerika Serikat dijelaskan pada Gambar 4.38 berikut.



Gambar 4.38 *Confusion Matrix* Pelatihan Keempat Amerika Serikat

Hasil dari percobaan pelatihan pada data Amerika Serikat dengan menggunakan *fine-tuning* berbeda tersebut ternyata memiliki hasil yang bervariatif. Hasil dari percobaan memiliki kelebihan dan kekurangan tersendiri pada setiap model. Hasil dari pelatihan tersebut disimpulkan dalam Tabel 4.19 berikut.

Tabel 4.19 Hasil Eksperimen Data Amerika Serikat

Eksp.	Kelas	*Precision*	*Recall*	*F1-score*	*Accuracy*	*Valid Loss*	*Over-fitting*	Catatan
Standar	*Negative*	0.73	0.76	0.74	0.79	Stabil	Tidak	Distribusi prediksi seimbang
	Positive	0.83	0.81	0.82				
Weighted Trainer Otomatis	*Negative*	0.87	0.51	0.64	0.75	Naik	Ada	Dominan ke kelas positif
	Positive	0.73	0.93	0.82				
Weighted Trainer Manual, *Threshold* 0.6	*Negative*	0.75	0.67	0.71	0.78	Naik	Ada	*Train loss* turun, *valid loss* naik
	Positive	0.79	0.85	0.82				

Eksp.	Kelas	*Precision*	*Recall*	*F1-score*	*Accuracy*	*Valid Loss*	*Over-fitting*	Catatan
Weighted Trainer Manual, *Threshold* 0.5	*Negative*	0.75	0.67	0.71	0.78	Fluktu-atif	Ada	Kinerja stabil meski *over-fitting*
	Positive	0.79	0.85	0.82				

Berdasarkan hasil evaluasi terhadap 4 model yang telah dilatih menggunakan dataset Amerika Serikat, maka disimpulkan bahwa model pertama memiliki performa yang baik secara keseluruhan. Hal ini ditunjukkan hasil *training* dengan akurasi sebesar 77%, *dengan precision, recall, dan F1-score* keseluruhan masing-masing 79% pada data uji. Dari hasil *confusion matrix* bahwa distribusi prediksi cukup seimbang dengan 62 prediksi kelas negatif benar dan 95 prediksi benar untuk kelas positif. Selain itu, kesalahan klasifikasi relatif kecil sebanyak 20 *false positive,* dan 23 *false negative.* Kemudian grafik evaluasi juga menunjukkan nilai *train loss* dan *validation loss* sama-sama menurun secara stabil tanpa ada lonjakan tajam, sehingga mendakan model tidak ada *overfitting.*

c) Perbandingan Hasil Evaluasi Indonesia dan Amerika Serikat

Hasil analisis pada eksperimen yang dilakukan pada data Indonesia dan Amerika Serikat memiliki perbedaan dalam hasil dari evaluasi model-model eksperimen pada data uji di setiap data. Hal ini dimana, hasil dari evaluasi model pada data Indonesia cenderung memiliki hasil akurasi, *recall, precision* dan *F1-score* memiliki nilai rata-rata dalam rentang 80 sampai 90. Berbeda dengan hasil dari evaluasi model pada data Amerika Serikat yang memiliki nilai rata-rata akurasi, *precison, recall* dan *F1-score* dengan rentang nilai 60 sampai 80. Hal ini terjadi karena perbedaan karakterisitik pada kedua bahasa berbeda, seperti gaya penulisan, penggunaan *slang*, idiom, atau ekspresi sentimen berbeda antar bahasa atau budaya. Dimana bahasa Indonesia dan bahasa Inggris memiliki perbedaan dalam tipologis yang cukup jauh, baik dalam struktur morfologi, sintaksis, maupun ekspresi sentimen. Seperti halnya yang disampaikan Ningyu Xu, et.al dalam penelitiannya menyatakan bahwa performa *zero-shot cross-lingual transfer* pada *multilangual* BERT sangat dipengaruhi oleh perbedaan sintaksis antar bahasa. Bahasa yang memiliki kesamaan tipologi cenderung menghasilkan transfer pengetahuan yang

lebih baik, sedangkan bahasa yang berbeda secara struktural akan menurun kinerja model [51].

Perbedaan karakteristik pada data Indonesia dan Amerika Serikat tersebut memiliki perbedaan gaya penulisan, penggunaan *slang,* idiom, atau ekspresi sentimen pada kedua data tersebut. Perbedaan karakteristik tersebut dapat dilihat pada Tabel 4.20 berikut.

Tabel 4.20 Perbedaan Karakteristik Penulisan Indonesia dan Amerika Serikat

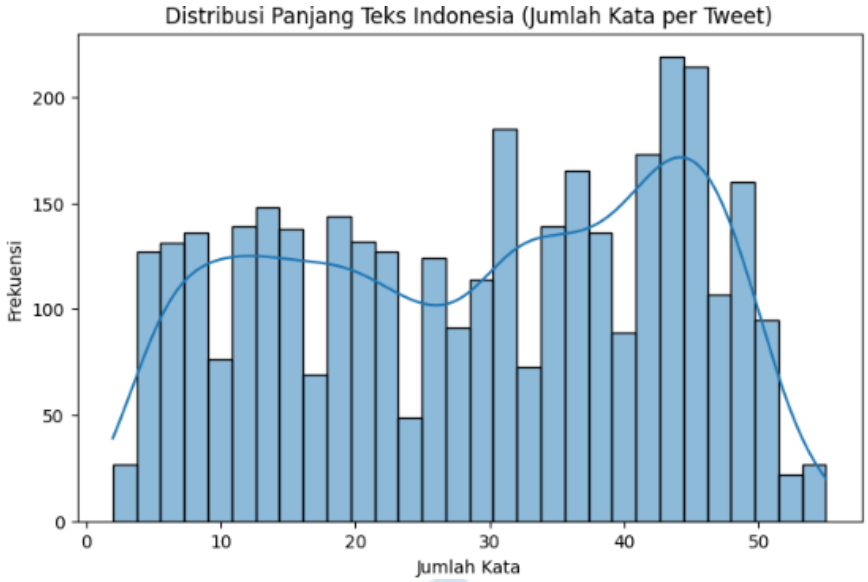
Indonesia		Amerika Serikat	
Teks	Sentimen	Teks	Sentimen
saya paham banget kadang perasaan takut atau ragu tentang pernikahan itu datang karena masih banyak hal yang patah kita alami atau raih dulu seperti healing dan menikmati waktu untuk diri sendiri nikah memang tanggung jawab besar dan enggak ada salahnya untuk fokus ke diri	Positif (mendukung refleksi diri)	a good marriage wont work with this ownership	Negatif (kritik hubungan pernikahan)
gue lebih milih nikah telat tp mental ama finansial mateng sih biar ntar pas udah nikah ga sedikit up story galau pusing mikirin rumah tangga takut bgt jir temen gue yg udah nikah saben hari ada aja yg disambatin lbh yg masih serumah sama mertua	Negatif (ragu/khawatir menikah)	marriage is not simply for people between the ages of to it requires full physical emotional and mental readiness age alone does not define marriage suitability a person must be physically prepared mentally mature and emotionally responsible to enter into such an	Positif (reflektif atau konstruktif pada pernikahan)

Pada kedua data di Tabel 4.20 tersebut bahwa data Indonesia memiliki gaya penulisan yang santai, spontan, dan bayak kalimat percakapan sehari-hari. Sedangkan pada data Amerika Serikat lebih formal, naratif, dan opini analitis. Kemudian pada data Indonesia memiliki banyak *slang* seperti kata “gue”, “ama”,

“ga” “jir”. Sedangkan pada data Amerika Serikat hampir tidak ada *slang* hanya lebih kepada bahasa standar. Selain itu pada data Indonesia memiliki ekspresi idiom seperti “*up story* galau”, “takut banget jir”, “*healing* dulu” yang langsung menunjukkan kepada emosi. Sedangkan pada data Amerika Serikat lebih figuratif atau metaforis seperti “*marriage won't work will this ownership*” yang berarti pernikahan yang sehat tidak bisa berjalan dengan sikap yang mengontrol pasangan.

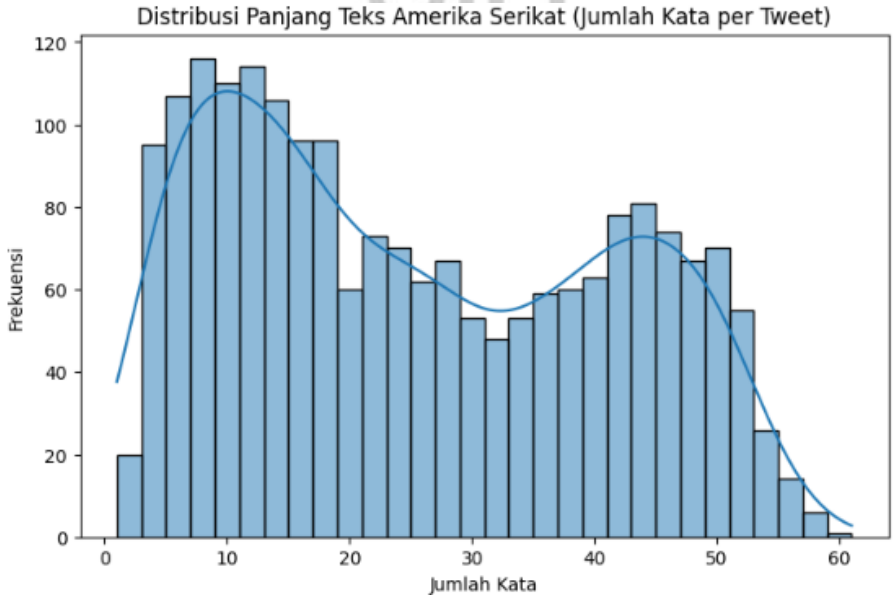
Selain itu, ekspresi sentimen pada data Indonesia lebih ekspresif dan eksplisit dengan langsung menyatakan kata “takut”, “galau”, “pusing”, dan “tertekan”. Sedangkan pada data Amerika Serikat lebih implisit. Kemudian pada data Indonesia fokus langsung kepada isu pribadi terkait takut menikah dan pengalaman sehari-hari. Sedangkan pada data Amerika Serikat fokus topik lebih luas membahas pernikahan dari aspek sosial, psikologis, bahkan politik. Akhirnya pada data Indonesia gaya bahasa lebih emosional dan eksplisit, banyak *slang* dan kata emosional, sehingga model lebih mudah mengenali sentimen dibandingkan data Amerika Serikat yang lebih formal, metaforis, dan implisit, sehingga model butuh pemahaman konteks lebih dalam, dan menjadikan performa relatif lebih rendah.

Kemudian pada data Indonesia, data memiliki distribusi teks yang merata. Dimana teks berada pada rentang mayoritas dari lebih dari 100 sampai lebih dari 200 karakter, sehingga relatif lebih panjang dan kaya konteks. Kemudian panjang teks lebih konsisten, sehingga model lebih mudah untuk menangkap pola sentimen karena tiap sampel mengandung cukup informasi. Distribusi panjang teks pada data Indonesia dijelaskan pada Gambar 4.39 berikut.



Gambar 4.39 Distribusi Panjang Teks Data Indonesia

Dibandingkan dengan data Indonesia, pada data Amerika Serikat panjang teks cenderung bimodal, yakni memiliki dua pundak besar dengan sekitar kurang dari 100 karakter dan lebih dari 250 karakter. Sehingga adanya teks yang sangat singkat dan kelompok teks yang lebih panjang. Dengan adanya hal ini dapat memengaruhi kinerja model, karena model harus mampu menangkap konteks baik dari teks singkat maupun teks panjang. Distribusi panjang teks pada data Amerika Serikat dijelaskan pada Gambar 4.40 berikut.



Gambar 4.40 Distribusi Panjang Teks Data Amerika Serikat

Maka dari itu, dengan adanya perbedaan pada karakteristik data, seperti gaya penulisan, *slang,* idiom, atau ekspresi sentimen antara kedua data pada data Indonesia dan Amerika Serikat dapat memberikan pengaruh dalam hasil performa pelatihan model. Hal ini menjadikan, meskipun kedua data dilatih dengan *pretrained model multilangual* BERT yang sama, akan tetapi hasil dari pelatihan pada kedua data dapat berbeda.

4. 3. Hasil Serta Pembahasan Rumusan Masalah Penelitian 3

Pelatihan model terbaik bertujuan untuk mengetahui hasil dari analisis sentimnen serta visualisasi terhadap data sentimen publik di media sosial X terhadap tren *"marriage is scary"*. Proses analisis sentimen dilakukan dengan memanfaatkan hasil pelatihan model terbaik untuk dapat mengklasifikasikan sentimen opini publik terkai "*marriage is scary*". Visualisasi pada tahap ini menggambarkan distribusi sentimen berdasarkan hasil sentimen opini publik yang muncul.

4.3.1. Hasil Analisis Sentimen Publik pada Media Sosial X Mengenai "*Marriage is Scary"*

Hasil dari penelitian analisis sentimen publik pada media sosial X terhadap isu "*marriage is scacy"* dengan pendekatan klasifikasi sentimen yang memanfaatkan arsitektur BERT. Data yang dianalisis berasal dari dua negara, yakni Indonesia serta Amerika Serikat. Hasil dari klasifikasi analisis sentimen ini disajikan dalam bentuk visualisasi-visualisasi berikut.

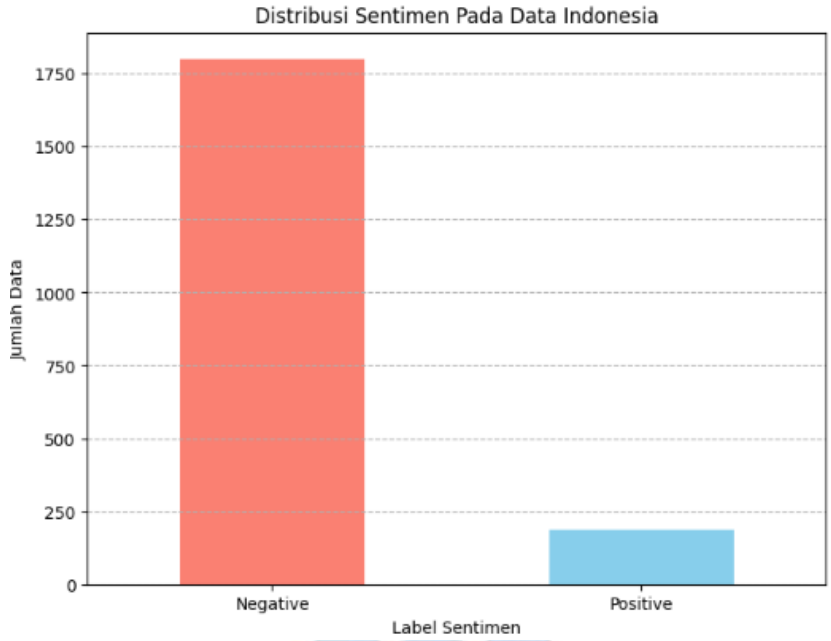
a) Hasil Analisis Sentimen Data Indonesia

Analisis sentimen pada data Indonesia dilakukan dengan menggunakan model pelatihan terbaik, yakni model pada eksperimen kedua dengan *epoch* 5. Model kemudian digunakan untuk memprediksi *subset* data kedua yakni *subset* 1988 untuk diberikan *label* sentimen pada data tersebut. Sehingga data *subset* tersebut dapat digunakan dalam analisis sentimen pada data Indonesia.

1) Distribusi Sentimen pada Data Indonesia

Pada Gambar 4.41 menunjukkan bahwa distribusi sentimen yang diperoleh dari hasil klasifikasi sentimen oleh model terbaik yakni eksperimen dengan 5 *epoch*, terhadap data pengguna media sosial di Indonesia. Dari hasil klasifikasi ternyata ditemukan dominasi sentimen negatif yang kuat. Hasil menunjukkan bahwa

terdapat lebih dari 1750 data diklasifikasin ke dalam sentimen negatif, sedangkan hanya kurang dari 250 data yang termasuk ke dalam sentimen positif. Hasil perbandingan tersebut menunjukkan banyaknya opini publik yang muncul di media sosial Indonesia mengenai isu *"marriage is scary"* bernada negatif. Distribusi sentimen pada data Indonesia dijelaskan pada Gambar 4.41 berikut.



Gambar 4.41 Distribusi Sentimen Data Indonesia

Dari distribusi tersebut, sehingga dapat diindikasikan bahwa pengguna media sosial X di Indonesia memiliki pandangan yang cenderung cemas, pesimis, atau skeptis terhadap respon tren "*marriage is scary*", yakni pada institusi pernikahan. Sehingga hal ini dapat mencerminkan adanya keresahan, trauma, ketakutan terhadap komitmen, atau terhadap kondisi sosial ekonomi yang turut memengaruhi persepsi masyarakat terhadap pernikahan.

2) *WordCloud Sentiment*

Dalam mengenal lebih dalam dari hasil analisis sentimen terhadap data Indonesia, hasil distribusi data tersebut divisualisasikan data-data dalam bentuk *WordCloud. WordCloud* digunakan untuk menggambarkan frekuensi kata yang muncul pada masing-masing sentimen. Visualisasi ini membantu dalam melihat fokus dari pembahasan publik terhadap isu "*marriage is scary"* dalam ranah positif maupun negatif.

Hasil dari *WordCloud* pada sentimen negatif pada data Indonesia, menunjukkan hasil yang kuat dan dominasi. *WordCloud* sentimen negatif menggambarkan kata-kata yang dominan dalam teks yang diprediksi memiliki nada negatif terhadap pernikahan. Kata-kata ini cenderung menunjukkan ketakutan, keraguan, penolakan atau pengalaman traumatis terkait pernikahan. Visualisasi ini berguna dalam memahami alasan-alasan utama mengapa sebagian opini menganggap pernikahan sebagi sesuatu yang menakutkan. Hasil dari *WordCloud* negarif dilihat pada Gambar 4.42 berikut.

Gambar 4.42 *WordCloud* Sentimen Negatif Indonesia

Hasil dari *WordCloud* tersebut bisa diketahui bahwa terdapat kata-kata yang berkonotasi negatif. Kata negatif yang muncul tersebut memiliki konotasi yang menyatakan ekspresi negatif terhadap respons dari diskusi tren *marriage is scary* pada *platform* media sosial X. Kata-kata berkonotasi negatif tersebut seperti pada kata-kata pada Tabel 4.21 berikut.

Tabel 4.21 Kalimat Sentimen Negatif Data Indonesia

Kata.	Teks Sentimen
“takut salah”	iya pengennya sih satu untuk selamanya ya makannya aku ga mau buru buru nikah **takut salah** pasangan
“takut salah” “pasangan”	tp ttp aja marriage is scary sampe diumur skrg aku blm menikah krn **takut salah** memilih **pasangan** dan calon anak yg jd korban broken home aku lbh ga mau calon anakku hrs menanggung krn kebodohanku yg salah memilih pasangan setrauma itu krn uda melihat byk cth nyata
“salah pilih”	rill mungkin di umur segini blm dikasih nikah sama tuhan karena tau bisa jd kita **salah pilih** atau takut dptin yg gk sesuai sama kita
“salah pilih”	bener takut **salah pilih** karena nikah itu semur hidup

Selain itu, hasil dari *WordCloud* sentimen positif pada data Indonesia menampilkan kata-kata yang banyak digunakan dalam teks-teks yang diklasifikasikan sebagai sentimen positif. Visualisasi pada *WordCloud* positif ini mencerminkan bagaimana opini publik mengekspresikan dukungan, harapan, kebahagiaan, serta penerimaan terhadap pernikahan dalam diskusi terhadap tren *marriage is scary*. Visualisasi *WordCloud* pada sentimen Indonesia ditunjukkan oleh Gambar 4.43 berikut

Gambar 4.43 *WordCloud* Sentimen Positif Indonesia

Hasil dari *WordCloud* positif diketahui terdapat beberapa kata yang muncul sebagai merepresentasikan sentimen positif. Kata-kata tersebut antara lain pada Tabel 4.22 berikut.

Tabel 4.22 Kalimat Sentimen Positif Data Indonesia

Kata	Teks Sentimen
“alhamdulillah”	**alhamdulillah** kalo dapat suami setia support dan baik semoga semua wanita baik dapat suami baik kalo takut zonk artinya harus lebih selektif cari cara dan berjuang menghindari dapat suami zonk bukan malah menghindari nikah dan atau menolak menjadi ibu
“suamiku” “alhamdulillah”	ih selisihnya mirip aku dan **suamiku** pas nikah paksu dan aku tadinya dia juga kaget karena masaaa kita sejauh ini jaraknya usia ya takut diledekin ceunah trus ywdah karena emang udh niat serius sih moo gimana lagi kita nikah nder punya anak **alhamdulillah**
“pasangan” “bahagia”	asik kak masalah dll itu ujian biar rumah tangga kita berwarna usia pernikahan lama pun bisa kena masalah tp bisa di lalui atau gak tergantung komunikasi sama **pasangan** kak aku awalnya jg takut tp setelah nikah jd tau banyak hal yg bikin **bahagia**

Berdasarkan hasil dari *WordCloud* positif di Indonesia, bahwa konteks bahasan positif masih tetap diutarakan oleh publik di platform sosial media X dalam melakukan diskusi terkait "*marriage is scary"*. Konteks bahasan positif terkait tren "*marriage is scary"* di Indonesia lebih cenderung mengarah kepada harapan, motivasi, dan ungkapan rasa syukur terhadap yang telah dilaksanakan oleh pengguna sosial media X. Akan tetapi pada isu ini di Indonesia sentimen negatif masih mendominasi opini publik terkait tren *"marriage is scary".*

Hasil dari analisis sentimen di data Indonesia menunjukkan adanya dominasi sentimen negatif dibandingkan dengan sentimen positif. Adanya dominasi sentimen negatif dibandingkan dengan sentimen positif di Indonesia ini, hal ini mennjukkan adanya keselarasan dengan fakta data realita. BPS menyatakan telah terjadi penurunan jumlah pernikahan pada tahun 2024 sebesar 69,75% pemuda di Indonesia belum menikah. Minat menikah mengalami penurunan dibandingkan dengan tahun 2023 sebesar 68,29% [48]. Dengan adanya fakta data tersebut, dengan hasil sentimen yang didominasi negatif pada opini pengguna platform media sosial X, sehingga disimpulkan bahwa pemuda pada saat ini mengalami penurunan minat terhadap pernikahan. Penurunan jumlah pernikahan disebabkan oleh beberapa faktor seperti adanya pergeseran pandangan terhadap pernikahan yang dipandang sebagai bukan prioritas utama yang harus dipilih, meningkatnya kemandirian pada perempuan, serta pemenuhan ekspektasi finansial pernikahan pada laki-laki [52].

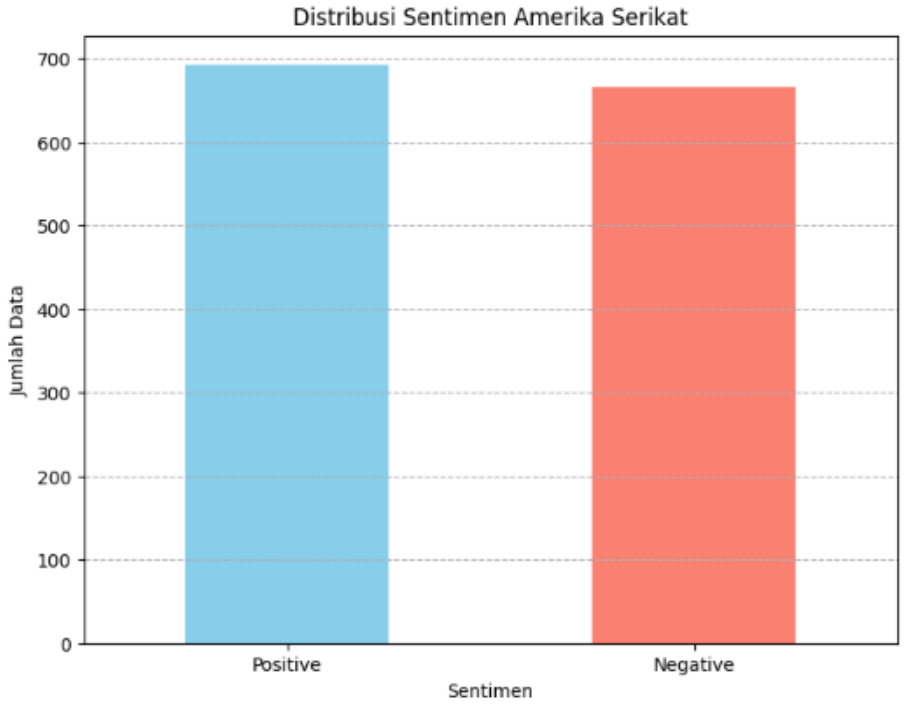
b) Hasil Analisis Sentimen Data Amerika Serikat

Proses analisis sentimen pada data Amerika serikat dilakukan dengan menggunakan model pelatihan terbaik yakni model pada percobaan pertama. Model tersebut kemudian digunakan untuk dilakukan prediksi pelabelan pada data *subset* 1359. Kemudian hasil prediksi tersebut dilakukan analisis sentimen lebih lanjut pada data Amerika serikat.

1) Distribusi Sentimen pada Data Sentimen Amerika Serikat

Berbeda dengan Indonesia, hasil distribusi sentimen pada data Amerika Serikat menunjukkan keseimbangan yang jauh lebih proporsional. Dari diagram batang, terdapat sekitar hampir dari 700 data positif dan lebih dari 600 data negatif. Hal ini artinya opini publik di Amerika Serikat terkait tren "*Marriage is Scary*" tidak

sepenuhnya pesimistis maupun optimistis. Masyarakat di Amerika Serikat cenderung terbagi dua dalam memaknai institusi pernikahan. Hasil tersebut ditunjukkan pada Gambar 4.44 berikut.



Gambar 4.44 Distribusi Sentimen Amerika Serikat

Keseimbangan distribusi ini dapat ditafsirkan sebagai refleksi dari keberagaman pengalaman dan perspektif budaya di Amerika Serikat. Meskipun sebagian masyarakat masih menyuarakan kekhawatiran terhadap pernikahan, tidak sedikit pula yang melihat pernikahan sebagai komitmen yang bermakna dan positif.

2) *WordCloud* Sentimen Amerika Serikat

WordCloud merupakan salah satu metode visualisasi yang menampilkan kata-kata yang paling sering muncul dalam kumpulan teks. Hasil analisis sentimen pada data Amerika Serikat ini memanfaatkan *WordCloud* untuk mengetahui kata-kata yang muncul dalam data sentimen, sehingga membantu dalam menganalisis ekspresi sentimen yang terkandung dalam opini publik terkait isu "*Marriage is Scary*" pada media sosial X di Amerika Serikat. Visualisasi ini memperkuat hasil klasifikasi sentimen yang telah dilakukan oleh model BERT terbaik, dan memberikan gambaran umum tentang persepsi masyarakat terhadap isu tersebut.

Hasil dari *WordCloud* untuk sentimen negatif memperlihatkan kata-kata berkonotasi negatif. Kata-kata ini menunjukkan adanya kekhawatiran, trauma, atau

pandangan skeptis terhadap institusi pernikahan. Hal ini mencerminkan pandangan sebagian masyarakat yang merasa bahwa pernikahan bisa membatasi kebebasan, membawa tekanan, atau berujung pada kegagalan seperti perceraian. Hasil dari *WordCloud* sentimen negatif pada data Amerika Serikat dijelaskan pada Gambar 4.45 berikut.

Gambar 4.45 *WordCloud* Sentimen Negatif Amerika Serikat

Hasil dari *WordCloud* pada Gambar 4.43 tersebut diketahui bahwa terdapat kata-kata yang berkonotasi negatif. Kata negatif yang muncul menyatakan ekpresi negatif seperti pada kata-kata Tabel 4.23 berikut.

Tabel 4.23 Kalimat Sentimen Negatif Data Amerika Serikat

Kata	Teks Sentimen
"wrong"	*yes a marriage with **wrong** person can become a curse but not every marriage*
"cheating"	*my to early just turned were full of up and downs and i was in a semi loveless marriage that ended with him **cheating** and a divorce it has definitely been hard but i have learned to surround myself with people who truly care about me and that its ok to put myself first*

Ungkapan sentimen positif terhadap konteks diskusi terkait tren "*marriage is scary"* pada Amerika Serikat menunjukkan kata-kata ungkapan pengalaman yang diutarakan terkait diskusi tersebut. Sehingga hasil dari *WordCloud* sentimen positif di data Amerika serikat dapat dilihat dalam Gambar 4. 46 berikut.

Gambar 4.46 *WordCloud* Sentimen Positif Amerika Serikat

Berdasarkan hasil klasifikasi model, beberapa teks opini publik dari masyarakat Amerika Serikat dikategorikan sebagai sentimen positif. Kata-kata yang muncul dominan dalam *WordCloud* sentimen positif seperti *"happy", "relashionship",* dan "*together*". Kemunculan kata-kata ini menggambarkan adanya ekspresi kasih sayang, kepercayaan terhadap institusi pernikahan, dan harapan akan hubungan yang harmonis. Contoh dari *WordCloud* tersebut seperti Tabel 4.24 berikut.

Tabel 4.24 Kalimat Sentimen Positif Data Amerika Serikat

Kata	Teks Sentimen
"happy" "marriage"	*but a lil bit of my heart wish **happy marriage** life would happen to me and i could be the bestfriendmother for my children*
"relationship" "together"	*yeah thats true they will change after marriage but havent they changed before in the **relationship** they might have its not about being same in the relationship having same things to do its always about trust that however the situations are they will end up **together***

Berdasarkan hasil analisis sentimen opini publik terkait isu "*Marriage is Scary*" di media sosial X, ditemukan adanya perbedaan pola persepsi antara masyarakat Indonesia dan Amerika Serikat. Mayoritas opini dari pengguna Indonesia menunjukkan sentimen negatif, dengan kata-kata berupa unkapan rasa takut, khawatir, dan beban mendominasi yang mencerminkan kekhawatiran terhadapa pernikahan.

Sebaliknya, opini publik di Amerika Serikat lebih imbang dengan kemunculan kata-kata positif seperti *"happy", "relashionship",* dan "*together*" di satu sisi, namun juga diimbangi oleh sentimen negatif seperti *wrong* dan *cheating*. Hal ini menunjukkan bahwa meskipun kekhawatiran terhadap pernikahan juga muncul di Amerika Serikat, akan tetapi tetap terdapat narasi positif yang cukup kuat mengenai harapan terhadap pernikahan yang sehat dan berbasis cinta.

Hasil analisis pada data Amerika Serikat juga menunjukkan adanya kecenderungan keseimbangan sentimen negatif dan positif. Meskipun data dari *Centers for Disease Control and Prevention* (CDC) Amerika Serikat yang menyatakan bahwa tingkat pernikahan di Amerika Serikat mengalami penurunan dalam dua dekade terakhir, dari tahun 200 sebesar 8,2 per 1000 populasi menurun menjadi 6,2 per 1000 populasi pada tahun 2022 [53], hal ini tidak serta-merta mencerminkan adanya penolakan terhadap institusi pernikahan itu sendiri. Akan tetapi menurut Survei *Pew Research Center* pada tahun 2023 sebanyak 73% pemuda di Amerika Serikat tidak merasakan tekanan untuk menikah dari orang tua, serta sebanyak 69% orang dewasa berusia 18 hingga 34 tahun mengatakan ingin menikah suatu saat nanti, meskipun cenderung menunda dan tidak menganggap pernikahan sebagia faktor utama untuk menjalankan kehidupan yang memuaskan [50]. Dengan demikian, meskipun tren pernikahan mengalami penurunan, akan tetapi opini publik terhadap pernikahan di media sosial cenderung netral, dimana disatu sisi memandang pernikahan sebagai tujuan jangka panjang, namun disisi lain, memendam kekhawatiran atau ketakutan terhadap pernikahan, seperti yang telah tercermin dalam hasil analisis sentimen yang seimabang antara positif dan negatif.

Dari hasi analisis sentimen pada kedua negara antara Indonesia dan Amerika serikat, dengan demikian dapat disimpulkan bahwa masyarakat Indonesia cenderung lebih pesimis terhadap pernikahan dibandingkan masyarakat Amerika Serikat yang memiliki pandangan lebih kompleks dan terbuka terhadap dinamika pernikahan. Sehingga dari kedua negara tersebut adanya perbedaan tanggapan terhadap tren *"marriage is scary"*.