

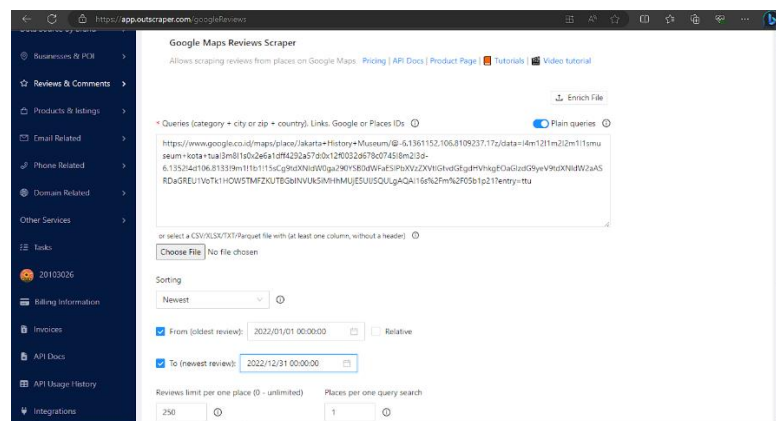
## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Hasil

##### 4.1.1 Pengumpulan Data

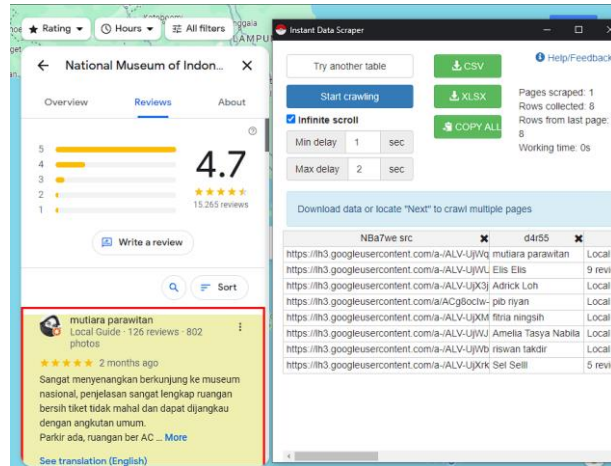
Penelitian ini menggunakan data ulasan yang diberikan pengunjung museum di Jakarta, antara lain *Jakarta History Museum*, Museum Nasional, Museum MACAN, Museum Bank Indonesia, dan Museum Wayang. Data ulasan yang digunakan penelitian ini adalah ulasan yang dibagikan pengunjung museum dalam kurun waktu lima tahun ke belakang, yaitu dari tahun 2018 hingga 2022. Data tersebut dikumpulkan melalui proses *scraping* pada *platform* Google Map Reviews menggunakan *tools* Outscraper dan Instant Data Scraper yang dilakukan pada tanggal 11 – 25 September 2023. Proses *scraping* menggunakan *tools* Outscraper diawali dengan membuka *website* dan *login*, kemudian pilih Google Maps Reviews Scraper, selanjutnya tempel *link* objek museum yang telah disalin pada kolom yang disediakan dan mengisi parameter *scraping*. Contoh proses *scraping* menggunakan *tools* Outscraper ditunjukkan pada Gambar 4.1.



Gambar 4. 1. Proses *scraping data* menggunakan *tools* Outscraper

Sedangkan, proses *scraping* menggunakan *tools* Instant Data Scraper diawali dengan *install extension* Instant Data Scraper pada Google Chrome, selanjutnya membuka ulasan objek museum pada platform Google Maps,

kemudian aktifkan *extensions*. Contoh proses *scraping* menggunakan *tools* Instant Data Scraper ditunjukkan pada Gambar 4.2.



Gambar 4. 2. Proses *scraping data* menggunakan *tools* Instant Data Scraper

Setelah dilakukan *scraping* menggunakan kedua *tools*, dilakukan penggabungan *dataset* dari beberapa *dataset* hasil *scraping* menjadi satu *dataset*. Hasil *scraping* menggunakan Outscraper dapat dilihat pada Gambar 4.3, sedangkan hasil *scraping* menggunakan Instant Data Scraper dapat dilihat pada Gambar 4.4

	name	review_text	year
0	Museum Sejarah Jakarta	Paling enak ngajak bocil kesini, karna ruang t...	2022
1	Museum Sejarah Jakarta	Benchmark kota Jakarta, seakan masuk mesin wak...	2022
2	Museum Sejarah Jakarta	Belajar sejarah	2022
3	Museum Sejarah Jakarta	Belajar sejarah tentang Jakarta. Akhir pekan r...	2022
4	Museum Sejarah Jakarta	Tempat yang asik buat bersantai bersama keluar...	2022

Gambar 4. 3. Contoh hasil *scraping* dengan Outscraper

	rsqaWe	wil7pd
0	sebulan lalu	Kota tua emg asik,\nTapi yang gak asik pedagan...
1	2 minggu lalu	Musium sejarah jakarta cocok untuk mengajak an...
2	2 bulan lalu	Anda bisa bernostalgia atau melihat kenangan m...
3	sebulan lalu	Bagus, terdapat barang lama yang dipajang, ser...
4	sebulan lalu	Klo ke sni mnding cari parkir an tiket dah,,ben...

Gambar 4. 4. Contoh hasil *scraping* menggunakan Instant Data Scraper

Berdasarkan Gambar 4.4 hasil *scraping* dengan Instant Data Scraper memiliki penamaan atribut dan format waktu yang berbeda dibandingkan yang menggunakan Outscraper. Oleh karena itu, dilakukan proses penyesuaian. Pertama, nama atribut mengikuti penamaan menggunakan Outscraper. Atribut “*rsqaWe*” menjadi “*year*” dan “*wil7pd*” menjadi “*review\_text*”. Kedua, penambahan atribut “*name\_value*” untuk menyimpan nama objek museum. Terakhir, mengubah value pada kolom “*year*” dengan kriteria: “2 tahun lalu” menjadi “2021”, “3 tahun lalu” menjadi “2020”, dst. Proses ini dapat dilihat pada Lampiran 1. Total *dataset* yang berhasil dikumpulkan di antaranya *Jakarta History Museum* sebanyak 1381 ulasan, *Museum Nasional* sebanyak 1015 ulasan, *Museum Bank Indonesia* sebanyak 914 ulasan, *Museum MACAN* sebanyak 1207 ulasan, dan *Museum Wayang* sebanyak 1034 ulasan. *Dataset* tersebut disimpan dengan format .csv. Contoh hasil *scraping* yang telah digabungkan dan dilakukan penyesuaian ditunjukkan pada Tabel 4.1.

Tabel 4. 1. Hasil *scraping data*

<i>name</i>	<i>review_text</i>	<i>year</i>
Museum Macan ( <i>Modern and Contemporary Art in Nusantara</i> )	Tempatnya super keren. Enak ajak anak2 kesini	2022

<i>name</i>		<i>review_text</i>	<i>year</i>
Museum Sejarah Jakarta		Tempat yang asik buat bersantai bersama keluarga dan sanak saudara, banyak museum yang bisa dikunjungi dan juga penampilan musisi dikawasan kota tua	2022
Museum Bank Indonesia		Mengenal lebih jauh tentang perkembangan perekonomian dan salah satu sejarah Indonesia, juga menambah wawasan kita.	2018
Museum Nasional Indonesia		Acara pameran yang sangat menginspirasi, banyak lukisan yang indah dan nilai seni yang tinggi, suatu pameran seni budaya yang harus dipertahankan keberadaannya, semoga pameran di musuem nasional bisa rutin dilakukan.	2020
Museum Wayang Jakarta		Sewaktu Jaman Penjajahan Belanda	2022

<i>name</i>	<i>review_text</i>	<i>year</i>
	Sebagai gerejanya orang-orang Belanda, bahkan di belakang ada makam Gubernur jaman Belanda JP COEN sekarang di alih Fungsi kan sebagai museum wayang, banyak contoh wayang dari seluruh Nusantara bahkan dari luar negeri juga ada wayang Cina, Korea dan lain lain.	

*Dataset* hasil *scraping* yang ditunjukkan pada Tabel 4.1 terdiri atas tiga variable, yaitu variabel “*name*” yang menunjukkan nama objek museum, “*review\_text*” menunjukkan ulasan yang telah dibagikan pengunjung terhadap objek museum terkait, dan “*year*” yang menunjukkan keterangan waktu ulasan dibagikan.

#### 4.1.2 *Data Preprocessing*

Setelah mengumpulkan data yang digunakan, dilakukan *preprocessing* terhadap data tersebut. *Data preprocessing* adalah tahapan transformasi dari data mentah menjadi data yang siap digunakan untuk dianalisis. Pada penelitian ini hanya menggunakan variabel *review\_text* dari data hasil *scraping*. Tahapan *preprocessing* yang dilakukan sebagai berikut.

##### 1) *Case folding*

Proses *case folding* bertujuan untuk menyamaratakan seluruh karakter pada *review\_text*. Proses ini dilakukan menggunakan metode `.lower()` pada objek string dalam Python yang ditunjukkan pada Lampiran 2.

Metode tersebut digunakan untuk mengonversi semua karakter menjadi huruf kecil atau *lowercase*. Contoh ulasan hasil *scraping* yang telah dilakukan proses *case folding* menggunakan Python ditunjukkan pada Tabel 4.2.

Tabel 4. 2. Hasil *case folding* dengan Python

Teks Ulasan	Hasil <i>Case folding</i>
Tempatnya super keren. Enak ajak anak2 kesini.	tempatnya super keren. enak ajak anak2 kesini.
Museum saat ini sangat menarik untuk dikunjungi. Anda dapat membawa teman, keluarga, dan menghabiskan 2-4 jam hanya untuk melihat berbagai seni, berbagai gaya dan bentuk dan juga Anda dapat membaca info dari seni tersebut dan mendapatkan makna di balik proses dan ide tersebut.	museum saat ini sangat menarik untuk dikunjungi. anda dapat membawa teman, keluarga, dan menghabiskan 2-4 jam hanya untuk melihat berbagai seni, berbagai gaya dan bentuk dan juga anda dapat membaca info dari seni tersebut dan mendapatkan makna di balik proses dan ide tersebut.
Keren... Tapi supaya gak ribet.. buat beli tiket masuk bisa dibeli e-commers perjalan seeperti taveloka atau tiket. Com	keren... tapi supaya gak ribet.. buat beli tiket masuk bisa dibeli e-commers perjalan seeperti taveloka atau tiket. com

## 2) *Convert emoticon*

Proses ini dilakukan konversi dari setiap emotikon yang terdapat di *review\_text* menjadi teks dengan menggunakan library `emoji`. Proses ini dilakukan menggunakan fungsi `demojize` dari library `emoji` untuk mengonversi emotikon dalam teks ulasan terkait yang ditunjukkan pada Lampiran 3. Dalam proses konversi emotikon dilakukan dalam bahasa

Indonesia. Contoh ulasan hasil *scraping* yang telah dilakukan proses *convert emoticon* menggunakan Python ditunjukkan pada Tabel 4.3.

Tabel 4. 3. Hasil *convert emoticon* dengan Python

Teks Ulasan	Hasil <i>Convert Emoticon</i>
kalau mau kesini enakan pilih slot dari pagi krn tutupnya jam 4 sore, kurang puas kalau sebentar 😊	kalau mau kesini enakan pilih slot dari pagi krn tutupnya jam 4 sore, kurang puas kalau sebentar :wajah_gembira_dengan_mata_bahagia:
memori indah sebelum masuk memori hp. selamat berkunjung! 😊	memori indah sebelum masuk memori hp. selamat berkunjung! :wajah_gembira:
meskipun jujur jauh dari rmh 😊	meskipun jujur jauh dari rmh :tertawa_terbahak-bahak:

Berdasarkan Tabel 4.3 dapat diketahui bahwa hasil *convert emoticon* berupa kata dalam Bahasa Indonesia yang merepresentasikan *emoticon* itu sendiri. Kata hasil konversi emotikon bisa mengandung emosi. Contohnya, kata “wajah gembira” yang mengandung emosi “bahagia”, “antisipasi”, dan “yakin” di kamus *NRC Emolex*. Contoh lain, kata “tertawa” yang mengandung emosi “bahagia” dan “terkejut”. Hasil konversi emotikon ini dapat membuat proses ekstraksi emosi menjadi lebih mendalam karena tidak hanya memperhatikan unsur teks saja tetapi juga emotikon yang terdapat pada teks ulasan.

### 3) *Remove punctuation and number*

Karakter angka, simbol, dan tanda baca tidak mengandung emosi apapun sehingga karakter-karakter tersebut perlu dihapuskan dari teks ulasan. Proses penghapusan ini dilakukan dengan menggunakan Python, dimana kode program yang digunakan untuk proses ini

ditunjukkan pada Lampiran 4. Proses ini diawali dengan menghapus tanda baca dari teks dengan menggunakan *library re* yang tersedia di Python. Kemudian, dilakukan normalisasi teks dengan menggunakan *library unicodedata*. *Library* tersebut bertujuan untuk mengubah karakter-karakter *unicode* ke dalam bentuk kompatibel dengan ASCII. Misalnya, karakter pangkat 2 (²) dapat diubah menjadi bentuk ASCII yang sesuai, yakni angka 2. Selanjutnya, menghapus karakter numerik atau angka baik negatif ataupun positif dengan menggunakan *library re*. Terakhir, menghapus *underscore* ( ) dari teks dan menggantikan dua atau lebih spasi berturut-turut dengan satu spasi. Contoh ulasan hasil *scraping* yang telah dilakukan proses *remove punctuation and number* menggunakan Python ditunjukkan pada Tabel 4.4.

Tabel 4. 4. Hasil *remove punctuation and number* dengan Python

Teks Ulasan	Hasil <i>Remove Punctuation and Number</i>
kalau mau kesini enakan pilih slot dari pagi krn tutupnya jam 4 sore, kurang puas kalau sebentar :wajah_gembira_dengan_mata_bahagia:	kalau mau kesini enakan pilih slot dari pagi krn tutupnya jam sore kurang puas kalau sebentar wajah gembira dengan mata bahagia
memori indah sebelum masuk memori hp. selamat berkunjung! :wajah_gembira:	memori indah sebelum masuk memori hp selamat berkunjung wajah gembira
meskipun jujur jauh dari rmh :tertawa_terbahak-bahak:	meskipun jujur jauh dari rmh tertawa terbahak bahak



#### 4) Convert slangword

*Slangword* merupakan kata-kata atau frasa-frasa yang sering digunakan dalam bentuk tidak formal atau digunakan dalam sehari-hari. Proses ini dilakukan untuk mengubah *slangword* menjadi bahasa baku yang sesuai dengan kaidah ejaan yang baik dan benar. Penelitian ini menggunakan kumpulan kosakata *slang* yang sudah ada dengan ada penambahan kosakata *slang* berdasarkan teks ulasan yang digunakan. Total kosakata *slang* pada kamus *slangword* yang digunakan sebanyak 16.574 kata. Kamus tersebut disimpan dalam format *.xlsm* yang ditunjukkan pada Tabel 4.5.

Tabel 4. 5. Contoh kosakata *slang*

Slang	Formal
woww	wow
aminn	amin
met	selamat
netaas	menetas
keberpa	keberapa
eeeehhh	eh
smga	semoga

Proses konversi kosakata *slang* dilakukan dengan menggunakan kode program yang dapat dilihat pada Lampiran 5. Penjelasan sintak program sebagai berikut.

a. `for i in range(len(df_jakarta_history_museum)):`  
 Berfungsi untuk mengiterasi melalui setiap baris (indeks).

b. `for word in df_jakarta_history_museum.review_text[i].split():`  
 Berfungsi untuk split setiap kata dalam kolom 'review\_text' pada baris ke-i.

c. `key_norm[key_norm['slang']] == word['formal'].values[0] if (key_norm['slang']) ==`

`word).any()` **else** `word` : Berfungsi untuk melakukan pencarian kata *slang*. Jika kata *slang* ditemukan maka kata *slang* digantikan dengan kata formal yang sesuai. Jika tidak ditemukan maka kata tersebut tetap tidak berubah.

- d. **if** `slangword: df_jakarta_history_museum.review_text[i]`  
`= slangword` : Berfungsi untuk memeriksa apakah variabel *slangword* yang menyimpan hasil tidak kosong. Jika tidak kosong berarti terdapat konversi *slangword* yang dilakukan, maka 'review\_text' pada baris ke-i diperbarui dengan hasil konversi *slangword*.

Contoh ulasan hasil *scraping* yang telah dilakukan proses *convert slangword* menggunakan Python ditunjukkan pada Tabel 4.6.

Tabel 4. 6. Hasil *convert slangword* dengan Python

<b>Teks Ulasan</b>	<b>Hasil Convert Slangword</b>
kalau mau kesini enakan pilih slot dari pagi krn tutupnya jam sore kurang puas kalau sebentar wajah gembira dengan mata bahagia	kalau mau kesini enakan pilih slot dari pagi karena tutupnya jam sore kurang puas kalau sebentar wajah gembira dengan mata bahagia
keren bgt tempatnya tp sedih pengunjung nya byk yg lebih mentingin foto drpd nikmatin seni nya jd nya pd berisik gitu	keren banget tempatnya tapi sedih pengunjung nya banyak yang lebih mentingin foto daripada menikmati seni nya jadi nya pada berisik begitu

##### 5) *Stopword removal*

*Stopword removal* adalah proses untuk menghapus kata-kata umum yang tidak memiliki makna. Proses ini menggunakan *library* Sastrawi dengan adanya penyesuaian. Pertama, daftar *stopword* yang terdapat di Sastrawi dicocokkan pada kamus *Emolex*, jika cocok

maka kata terkait akan dihapus dari daftar *stopword*. Hal ini bertujuan untuk mencegah menghapus kata-kata yang sebenarnya mengandung emosi. Kedua, dilakukan penambahan *stopword* yang belum tersedia di Sastrawi berdasarkan teks ulasan penelitian. Contohnya, kata “kalo”, “kak”, “nya”, “semoga”, “kali”, “ya”, “deh”, “sih”, “banget”, “pas”, “masuk”, “kesini”, dan “kesana”. Kode program proses ini dapat dilihat pada Lampiran 6. Penjelasan sintak program sebagai berikut.

- a. `stopwords_ind = stopwords.words('indonesian') :`  
 Berfungsi untuk *import* daftar *stopword* dalam Bahasa Indonesia kemudian disimpan pada variabel `stopwords_ind`.
- b. `file_path = '../Id-NRC-EmoLex.xlsx'`  
`remove_stopword = pd.read_excel(file_path) :` Berfungsi untuk *import* kosakata di kamus *Emolex* sebagai daftar *stopword* yang akan dihapus dari daftar *stopword* sebelumnya.
- c. `remove_stopword = remove_stopword[remove_stopword['total']>0] :` Berfungsi memfilter daftar *stopword* yang akan dihapus berdasarkan total emosi yang dikandung kata tersebut. Jika sebuah kata pada daftar *stopword* mengandung satu atau lebih emosi, maka kata tersebut akan dihilangkan pada variabel `remove_stopword`.
- d. `remove_stopword = remove_stopword['Indonesian Word'].tolist() :` Berfungsi untuk mengubah tipe data variabel `remove_stopword` menjadi *list*.
- e. `add_remove_stopword = ['tidak']`  
`remove_stopword = remove_stopword + add_remove_stopword :` Berfungsi untuk menambahkan *list* `remove_stopword`.
- f. `more_stopword = ['aargh', 'aawww', 'aduhai', ..., 'ya', 'yoi', 'yups']`  
`stopwords_ind = stopwords_ind + more_stopword :` Berfungsi menambahkan daftar *stopword* pada variabel `stopwords_ind`.

- g. `stopwords_ind = [x for x in stopwords_ind if x not in remove_stopword]` : Berfungsi untuk menghapus *stopword* yang terdapat di variabel `remove_stopword`.
- h. `for i in range(len(df_jakarta_history_museum)):` Berfungsi untuk mengiterasi melalui setiap baris (indeks).
- i. `clean_words = []` : Berfungsi sebagai variabel yang menyimpan kata-kata setelah dilakukan *stopword removal*.
- j. `text = df_jakarta_history_museum.review_text[i].split():` Berfungsi untuk memecah kalimat dalam kolom 'review\_text' pada baris ke-i menjadi kata-kata.
- k. `for word in text:` Berfungsi untuk melakukan iterasi melalui setiap kata dalam kalimat.
- l. `if word not in stopwords_ind:` Berfungsi untuk melakukan pemeriksaan apakah kata bukan termasuk ke dalam *stopword*. Jika bukan *stopword*, maka kata tersebut ditambahkan ke dalam 'clean\_words'.
- m. `df_jakarta_history_museum.review_text[i] = ".join(clean_words):` Berfungsi untuk menggabungkan kembali kata-kata yang tersisa dalam *list* 'clean\_words' menjadi satu string. Kemudian, *replace* nilai dalam kolom 'review\_text' pada baris ke-i.

Contoh ulasan hasil *scraping* yang telah dilakukan proses *stopword removal* menggunakan Python ditunjukkan pada Tabel 4.7.

Tabel 4. 7. Hasil *stopword removal* dengan Python

Teks Ulasan	Hasil <i>Stopword Removal</i>
karya seni yang ditampilkan bagus lebih ke karya seni tingkat tinggi jadi bagi orang awam seperti saya agak susah mencernanya petugasnya ramah dan baik	karya seni ditampilkan bagus karya seni tingkat tinggi orang awam susah mencernanya petugasnya ramah baik mengarahkan pengunjung tempatnya nyaman parkir gedungnya

<b>Teks Ulasan</b>	<b>Hasil <i>Stopword Removal</i></b>
mengarahkan pengunjung tempatnya nyaman walaupun parkir dan masuk gedungnya agak membingungkan tiketnya bisa pesan online atau beli on the spot di kisaran ribu untuk akhir pekan	membingungkan tiketnya pesan online beli on the spot kisaran ribu pekan
ini adalah galeri seni modern yang mengesankan pasti menempatkan pemandangan seni jakarta setara dengan kota kota seperti new york namun perlu diketahui bahwa sebagai galeri seni instalasinya terus berubah secara berkala jadi apa yang anda lihat di media sosial mungkin tidak ada	galeri seni modern mengesankan pasti menempatkan pemandangan seni jakarta setara kota kota new york galeri seni instalasinya berubah berkala lihat media sosial tidak

*Dataset* yang telah dilakukan *preprocessing* disimpan dalam format .csv dan contoh hasilnya dapat terlihat pada Tabel 4.8.

Tabel 4. 8. Hasil *data preprocessed*

<b>Teks Ulasan</b>	<b>Teks <i>Preprocessed</i></b>
Rapi bagus, dan karya <sup>2</sup> nya unik ..	rapi bagus karya unik
Lokasi nya enak, bersih, harga tiket lumayan murah, sayangnya saya merasa art nya kurang banyak yang bisa di explore ya , paling mentok sejam di sini juga udah gatau mau ngapain lagi.	lokasi enak bersih harga tiket lumayan murah sayangnya merasa art jelajah mentok sejam tau ngapain lagi pengen foto foto menghabiskan jam cukup orang berjualan artnya oke harga tiketnya murah tidak berharap

Teks Ulasan	Teks <i>Preprocessed</i>
<p>Buat yang pengen foto-foto saya rasa menghabiskan satu jam saja sudah cukup di sini.</p> <p>Terlalu banyak orang yang berjualan , artinya kurang banyak.</p> <p>Tapi oke lah, harga tiketnya kan murah juga . Jadi tidak bisa berharap banyak</p>	
<p>Maaf, tapi saya berharap lebih untuk museum ini..</p>	<p>maaf berharap museum</p>
<p>Harga sesuai dengan experience yg didapat</p>	<p>harga sesuai pengalaman</p>
<p>Haus abis nyoba segala pose? #AQUADULU</p>	<p>haus habis mencoba pose aquadulu</p>
<p>Kalau mau kesini enakan pilih slot dari pagi krn tutupnya jam 4 sore, kurang puas kalau sebentar 😊</p>	<p>enakan pilih slot pagi tutupnya jam sore puas sebentar wajah gembira mata bahagia</p>
<p>Ide dating?coba kunjungi Museum MACAN</p>	<p>ide berkencan coba kunjungi museum macan</p>
<p>Di Museum Macan, ada kafe dan toko souvenir. Kamar mandinya pun bersih. Di dekat pintu masuk, ada space untuk foto di infinity room (ruangan kecil dengan pantulan bola lampu). Waktu itu saya mesti</p>	<p>museum macan kafe toko souvenir kamar mandinya bersih dekat pintu space foto infinity ruangan ruangan pantulan bola lampu waktu mesti mengantri foto infinity ruangan orang dibatasi</p>

Teks Ulasan	Teks <i>Preprocessed</i>
ngantri untuk foto di infinity room. Setiap orang yang masuk dibatasi oleh waktu yang super singkat oleh penjaga. Tapi kita bisa masuk ke situ lagi kalau belum puas dengan hasil foto yang didapat.	waktu super singkat penjaga situ lagi puas hasil foto

#### 4.1.3 Ekstraksi Emosi dengan *NRC Emolex*

Tahapan selanjutnya adalah ekstraksi emosi yang bertujuan mengetahui emosi dari setiap ulasan dan memberikan label pada ulasan pengunjung museum. Ekstraksi emosi penelitian ini menggunakan *NRC Emotion Lexicon* Saif Mohammad berbahasa Indonesia. Kamus ini terdapat delapan emosi, yakni *anger* (marah), *anticipation* (antisipatif), *fear* (takut), *disgust* (jijik atau muak), *sadness* (sedih), *surprise* (terkejut), *joy* (bahagia), dan *trust* (yakin). Secara umum, proses ekstraksi emosi yang dilakukan adalah memberi nilai ‘1’ untuk setiap kata dalam kalimat yang mewakili emosi marah, antisipatif, takut, muak atau jijik, sedih, terkejut, bahagia, atau yakin berdasarkan yang ada di kamus *Emolex*. Apabila tidak ada emosi yang terwakilkan maka akan diberi nilai ‘0’.

Proses ekstraksi emosi dilakukan dengan men-*split* kalimat ulasan menjadi kata per kata. Dimana setiap kata tersebut akan ditentukan label emosi hingga seluruh kata sudah terdapat labelnya. Setelah itu, dilakukan perhitungan total frekuensi dari setiap emosi pada seluruh teks ulasan. Total frekuensi ini akan digunakan pada proses analisis emosi dari setiap objek museum. Contohnya, proses ekstraksi emosi dari ulasan “bagus rapi” diawali ulasan tersebut di-*split* menjadi kata “bagus” dan “rapi”. Kemudian, setiap kata tersebut akan diberikan label sesuai emosinya dengan diberi nilai

‘1’ atau ‘0’ pada emosi terkait. Terakhir, dilakukan pejumlahan setiap emosi dari masing-masing kata. Proses pelabelan dari ulasan tersebut ditunjukkan pada Tabel 4.9.

Tabel 4. 9. Contoh proses ekstraksi emosi pada kalimat ulasan

Ulasan	<i>Anger</i>	<i>Anticipation</i>	<i>Disgust</i>	<i>Fear</i>	<i>Joy</i>	<i>Sadness</i>	<i>Surprise</i>	<i>Trust</i>
bagus	0	1	0	0	1	0	1	1
rapi	0	0	0	0	0	0	0	0
bagus rapi	0	1	0	0	1	0	1	1

Kode program yang digunakan untuk proses tersebut dapat dilihat pada Lampiran 7. Penjelasan sintak program sebagai berikut.

- a. `dictionary = pd.read_excel("../Id-NRC-EmoLex.xlsx"):`  
Berfungsi untuk *load NRC Emolex*.
- b. `anger=[]` : Berfungsi membuat *list* yang berisi kata-kata yang terikat dengan emosi tertentu.
- c. `for i in range(1, len(dictionary)):` Berfungsi untuk mengiterasi sepanjang "dictionary".
- d. `kata = dictionary.iloc[i,0]` : Berfungsi untuk mengambil value dalam kolom pertama dari baris saat ini dan menyimpan ke variabel "kata".
- e. `if(dictionary.iloc[i,1])==1: anger.append(kata)` : Berfungsi untuk memeriksa nilai dari kolom. Jika nilai sama dengan 1 maka kata tersebut ditambahkan ke *list* emosi yang sesuai. Artinya, kata mengandung emosi yang sesuai.
- f. `kamus_pattern_anger = r'\b(?:' + '|'.join(re.escape(k) for k in anger) + r')\b'` : Membuat ekspresi reguler dari kamus.
- g. `emosi = []` : Berfungsi membuat *list* emosi yang menyimpan hasil ekstraksi emosi.
- h. `for i in range(len(df_jakarta_history_museum)):` Berfungsi untuk melakukan iterasi sebanyak data ulasan.



- i. `baris = []` : Berfungsi untuk menyimpan hasil analisis emosi dari satu kalimat.
- j. `f_marah = 0` : Berfungsi untuk inisialisasi variabel yang menyimpan bobot emosi tertentu.
- k. `kalimat = df_jakarta_history_museum.review_text[i]`: Berfungsi untuk mengambil teks ulasan pada indeks ke-i dan menyimpannya ke dalam variabel "kalimat".
- l. `baris.append(kalimat)`: Berfungsi untuk menambahkan teks ulasan ke dalam *list*.
- m. `kata_marah_ditemukan = re.findall(kamus_pattern_anger, kalimat)`: Berfungsi untuk mencari kata dalam kalimat menggunakan `kamus_pattern` untuk setiap emosi.
- n. `if kata_marah_ditemukan: f_marah = len(kata_marah_ditemukan)`: Berfungsi untuk melakukan pemeriksaan apakah kata ditemukan atau tidak. Apabila ditemukan maka hitung jumlah kata yang ditemukan.
- o. `baris.append(f_marah)`: Befungsi untuk menambahkan jumlah kata yang ditemukan untuk setiap emosi ke dalam *list* baris.
- p. `emosi.append(baris)`: Berfungsi untuk menambahkan *list* baris ke dalam *list* emosi.

Contoh kalimat ulasan yang telah dilakukan ekstraksi emosi dapat dilihat pada Tabel 4.10. Dimana setelah semua ulasan diberikan label emosi maka dilakukan perhitungan total frekuensi untuk semua emosi dari *dataset*. Contohnya pada Tabel 4.10, total emosi antisipatif 9, takut 1, bahagia 8, sedih 2, terkejut dan yakin 5, serta emosi sisanya adalah 0.

Tabel 4. 10. Contoh hasil ekstraksi emosi pada kalimat ulasan

No.	Ulasan	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
1.	tempatnya super keren enak ajak anak	0	1	0	0	1	0	0	0
2.	tempatnya bagus	0	1	0	0	1	0	1	1

No.	Ulasan	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
3.	karya seni ditampilkan bagus karya seni tingkat tinggi orang awam susah mencernanya petugasnya ramah baik mengarahkan pengunjung tempatny nyaman parkir gedungnya membingungkan tiketnya pesan online beli on the spot kisaran ribu pekan	0	7	0	1	6	2	4	4
...	...	...	...	...	...	...	...	...	...
<i>n</i>	lukisan lukisan mengerti instagramable	0	0	0	0	0	0	0	0
<b>Total Emosi</b>		0	9	0	1	8	2	5	5

Berdasarkan hasil ekstraksi emosi dengan *NRC EmoLex*, dapat diketahui pada satu teks ulasan dapat mengandung beberapa emosi. Oleh karena itu, dilakukan pelabelan ke dalam kelas emosi dengan memilih emosi yang paling dominan yang dapat dilihat pada Lampiran 9. Penjelasan sintak program sebagai berikut.

- a. `emosi_new = []` : Berfungsi untuk membuat *list* `emosi_new` yang akan menampung teks ulasan beserta hasil pelabelan emosi.
- b. `for i in emosi:` Berfungsi untuk melakukan iterasi

- c. `baris=[]` : Berfungsi untuk membuat *list* `baris` yang akan menampung sementara hasil pelabelan emosi.
- d. `baris.append(i[0])` : Berfungsi menambahkan teks ulasan ke *list* `baris`.
- e. `max_value = max(i[1:])` : Berfungsi untuk mencari skor emosi tertinggi
- f. `if(max_value > 0)` : Berfungsi untuk memeriksa apakah skor tertinggi bukan sama dengan 0.
- g. `for j in range(1, 9)` : Berfungsi untuk melakukan iterasi sebanyak jumlah emosi.
- h. `bobot=i[j]/max_value` : Berfungsi untuk membagi skor dari setiap emosi dengan skor tertinggi.
- i. `if(bobot==1): baris.append(1)` : Jika hasil pembagian sama dengan 1 maka beri skor 1.
- j. `else: baris.append(0)` : Jika hasil pembagian tidak sama dengan 1 maka beri skor 0.
- k. `else: for j in range(1, 9): baris.append(0)` : Apabila `max_value` lebih kecil atau sama dengan 0 maka beri skor 0 untuk semua emosi.
- l. `emosi_new.append(baris)` : Berfungsi menambahkan list `baris` ke list `emosi_new`.

Tahapan ini diawali dengan memilih emosi yang memiliki skor paling tinggi, kemudian membagi *score* setiap emosi dengan *score* emosi paling tinggi. Apabila hasil pembagian  $< 1$  maka emosi tersebut diberi *score* '0', sedangkan apabila hasil pembagian  $= 1$  maka emosi tersebut diberi *score* '1'. Dari proses tersebut menghasilkan dua kelas tambahan, yaitu *mix-emotion* dan *no-emotion*. *Mix-emotion* artinya data ulasan yang memiliki lebih dari satu emosi dan *no-emotion* artinya data ulasan yang diidentifikasi tidak mengandung emosi. Contoh hasil pelabelan emosi ditunjukkan pada Tabel 4.11.

Tabel 4. 11. Hasil pelabelan emosi

Ulasan	<i>Anger</i>	<i>Anticipation</i>	<i>Disgust</i>	<i>Fear</i>	<i>Joy</i>	<i>Sadness</i>	<i>Surprise</i>	<i>Trust</i>	<i>Label</i>
tempatnya super keren enak ajak anak	0	1	0	0	1	0	0	0	<i>Mix-emotion</i>

Ulasan	<i>Anger</i>	<i>Anticipation</i>	<i>Disgust</i>	<i>Fear</i>	<i>Joy</i>	<i>Sadness</i>	<i>Surprise</i>	<i>Trust</i>	<i>Label</i>
karya seni ditampilkan bagus karya seni tingkat tinggi orang awam susah mencernanya petugasnya ramah baik mengarahkan pengunjung tempatny nyaman parkir gedungnya membingungkan tiketnya pesan online beli on the spot kisaran ribu pekan	0	1	0	0	0	0	0	0	<i>Anticipation</i>
lukisan lukisan mengerti instagramable	0	0	0	0	0	0	0	0	<i>No-emotion</i>

#### 4.1.4 Klasifikasi dengan SVM

*Dataset* dibagi menjadi *data training* dan *data testing* dengan proporsi masing-masing 70% dan 30%. Dengan *review\_text* sebagai data fitur dan *label* sebagai data target. Data dilakukan fitur ekstraksi atau pembobotan dengan menggunakan TF-IDF. Proses pembobotan ini menggunakan `library TfidfVectorizer()` dengan `ngram_range=(1,1)` yang tersedia di Python yang ditunjukkan pada Lampiran 10. Penjelasan sintak penggunaan TF-IDF.

- a. `tf_idf_1 = TfidfVectorizer(ngram_range=(1,1))` : Membuat objek `tf_idf_1` yang menerapkan metode TF-IDF dengan hanya mempertimbangkan *unigram* (kata tunggal).
- b. `tf_idf_1.fit(new_df_museum['review_text'])` : Berfungsi melakukan *fitting* model TF-IDF ke data teks ulasan yang terdapat di kolom `review_text`.
- c. `data_tf_idf_1 = pd.DataFrame(X_tf_idf_1, columns=tf_idf_1.get_feature_names_out())` : Berfungsi membuat *dataframe* baru yang berisi representasi TF-IDF dari teks.
- d. `X_tfidf_1 = np.array(data_tf_idf_1)` : Berfungsi untuk mengubah tipe menjadi array numpy.
- e. `y_tfidf_1 = np.array(new_df_museum['label'])` : kolom label akan menjadi data *y* (target).
- f. `X_train, X_test, y_train, y_test = train_test_split(X_tfidf_1, y_tfidf_1, test_size=0.3, random_state=42)` : Berfungsi membagi data menjadi data *training* dan *testing* dengan perbandingan 70:30.

Penjelasan sintak penggunaan *chi-squared*.

- a. `chi2_features = SelectKBest(chi2, k=3000)` : Penggunaan seleksi fitur *k-best* untuk memilih top 3000 fitur berdasarkan skor *chi-squared*. Jumlah fitur yang dipilih berdasar sekitar setengah dari jumlah fitur original.
- b. `X_kbest_features = chi2_features.fit_transform(X_tfidf_1, y_tfidf_1)` : Berfungsi melakukan transformasi fitur pada data.

Penjelasan sintak penggunaan PCA

- a. `pca = PCA(n_components=3000)` : Penggunaan PCA untuk mereduksi dimensi data sesuai `n_components` dalam sintak ini dipilih 3000 komponen. Nilai `n_components` ditentukan berdasar persentase varians.

- b. `X_train_pca = pca.fit_transform(X_train)` : Mengaplikasikan PCA pada data pelatihan.
- c. `X_test_pca = pca.transform(X_test)` : Mengaplikasikan PCA pada data uji.

#### Penjelasan sintak SMOTE:

- a. `smote = SMOTE(random_state=42)` : Membuat objek SMOTE dengan `random_state=42`.
- b. `X_train_smote, y_train_smote = smote.fit_resample(data_train_tf_idf_1, y_train)` : Proses *resampling* dengan SMOTE pada data *training*.
- c. `X_train_smote_arr = np.array(X_train_smote)`: Konversi objek data *training* yang telah di-*resampling* menjadi array.
- d. `y_train_smote_arr = np.array(y_train_smote)` : Konversi objek target label yang sesuai dengan data *training* yang telah di-*resampling* menjadi array.

#### Penjelasan sintak pemodelan

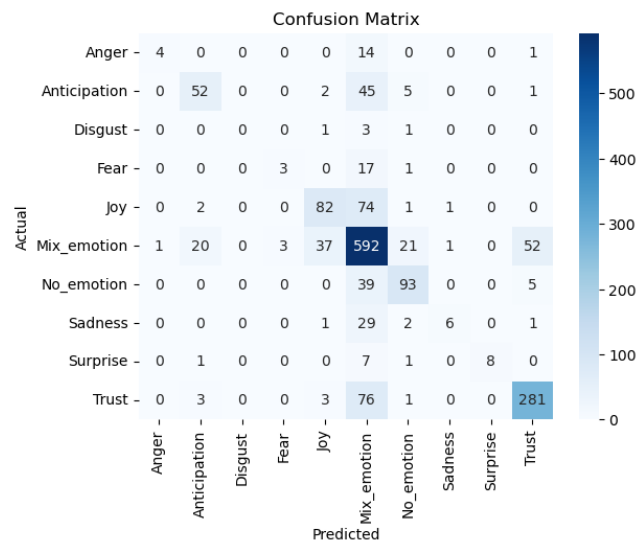
- a. `model = SVC(kernel='linear', C=5)` : Membuat model *support vector machine classifier* dengan `kernel='linear'` dan `C=5`.
- b. `model.fit(X_train, y_train)` : Melatih model dengan data *training*.
- c. `y_pred = model.predict(X_test)` : Melakukan prediksi pada data uji.
- d. `accuracy = accuracy_score(y_test, y_pred)` : Melakukan evaluasi performa model dengan akurasi
- e. `precision = precision_score(y_test, y_pred, average='macro')` : Melakukan evaluasi performa model dengan *precision*.
- f. `recall = recall_score(y_test, y_pred, average='macro')` : Melakukan evaluasi performa model dengan *recall*.
- g. `f1 = f1_score(y_test, y_pred, average='macro')` : Melakukan evaluasi performa model dengan *F1-score*.

- h. `print("Accuracy:", accuracy)` : Menampilkan hasil akurasi.
- i. `print("Precision:", precision)` : Menampilkan hasil presisi.
- j. `print("Recall:", recall)` : Menampilkan hasil *recall*.
- k. `print("F1-Score:", f1)` : Menampilkan hasil *F1-Score*.
- l. `print("\nClassification Report:\n", classification_report(y_test, y_pred))` : Menampilkan classification report.

Hasil pemodelan SVM dengan kernel linear dan  $C = 5$  sebagai berikut.

a. **Skenario 1: Non-Sampling, 10 Emosi, Seleksi Fitur**

*Dataset* tidak dilakukan *resampling* dengan menggunakan 10 emosi. Jumlah fitur awal 5826 dengan seleksi fitur *chi-squared* menjadi 3000 fitur. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.5 dapat dihitung hasil evaluasi model, yaitu *accuracy* 70,32%, *precision* 65,9%, *recall* 42,5%, dan *F1-score* 48,26%. Artinya, model berhasil 70,32% data dengan benar dari keseluruhan data, dimana hanya 65,9% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.

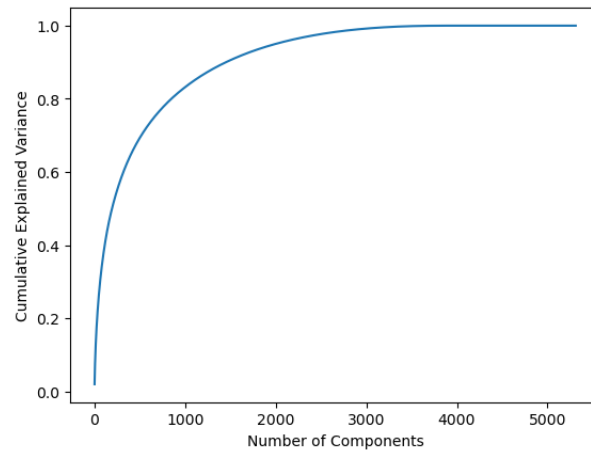


Gambar 4. 5. *Confusion matrix* skenario 1

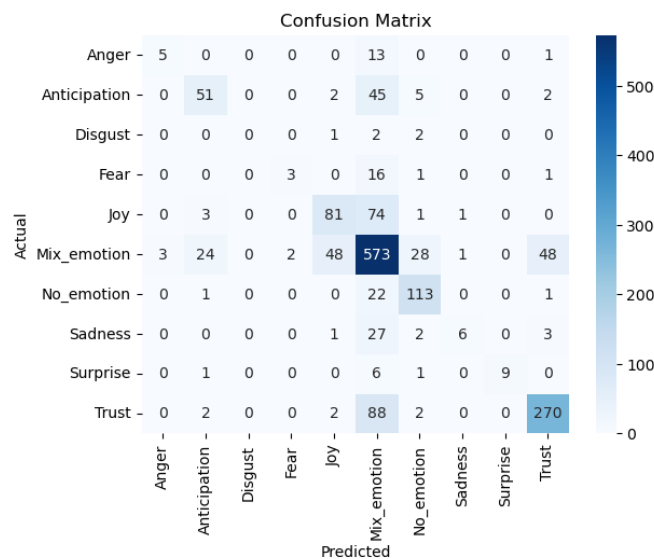
b. **Skenario 2: Non-Sampling, 10 Emosi, PCA**

*Dataset* tidak dilakukan *resampling* dengan menggunakan 10 emosi. Jumlah fitur awal 6899 dengan PCA menjadi 3000 fitur berdasarkan grafik yang ditunjukkan pada Gambar 4.6. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.7 dapat dihitung hasil evaluasi model, yaitu *accuracy* 69,69%, *precision* 64,15%, *recall* 44,35%, dan *F1-score* 49,19%. Artinya, model berhasil 69,69% data dengan benar dari keseluruhan data, dimana hanya 64,15% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.





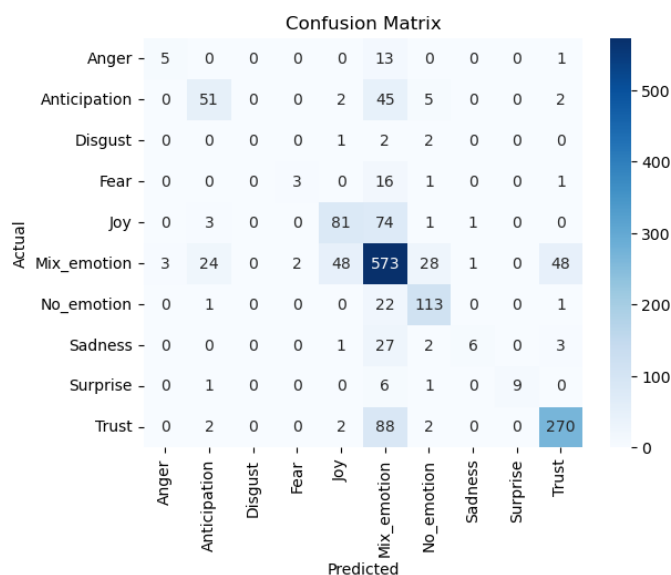
Gambar 4. 6. *Cumulative explained variance* pada non-sampling 10 emosi



Gambar 4. 7. *Confusion matrix* skenario 2

- c. **Skenario 3: Non-Sampling, 10 Emosi, Non seleksi fitur dan PCA**  
*Dataset* tidak dilakukan *resampling* dengan menggunakan 10 emosi. Jumlah fitur awal 5826 tanpa seleksi fitur dan PCA. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.8 dapat dihitung hasil evaluasi model, yaitu *accuracy* 69,69%, *precision* 64,15%, *recall* 44,35%, dan *F1-score* 49,19%. Artinya, model berhasil 69,69% data dengan benar dari keseluruhan data, dimana hanya 64,15% di antaranya

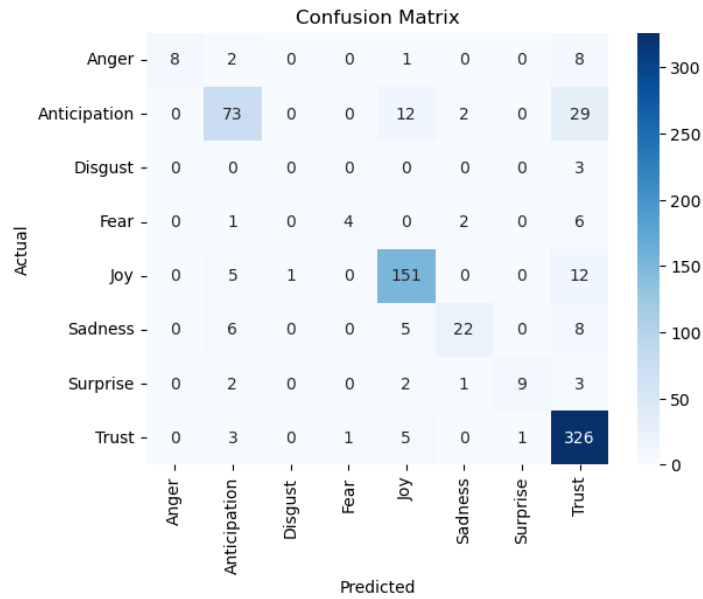
yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



Gambar 4. 8. *Confusion matrix* skenario 3

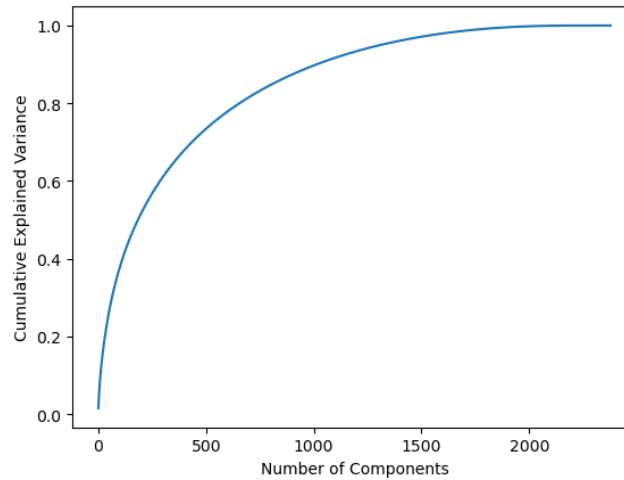
d. **Skenario 4: Non-Sampling, 8 Emosi, Seleksi Fitur**

*Dataset* tidak dilakukan *resampling* dengan menggunakan 8 emosi. Jumlah fitur awal 4414 dengan seleksi fitur menjadi 2000 fitur. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.9 dapat dihitung hasil evaluasi model, yaitu *accuracy* 83,05%, *precision* 74,8%, *recall* 53,59%, dan *F1-score* 60,25%. Artinya, model berhasil 83,05% data dengan benar dari keseluruhan data, dimana hanya 74,8% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



Gambar 4. 9. *Confusion matrix* skenario 4

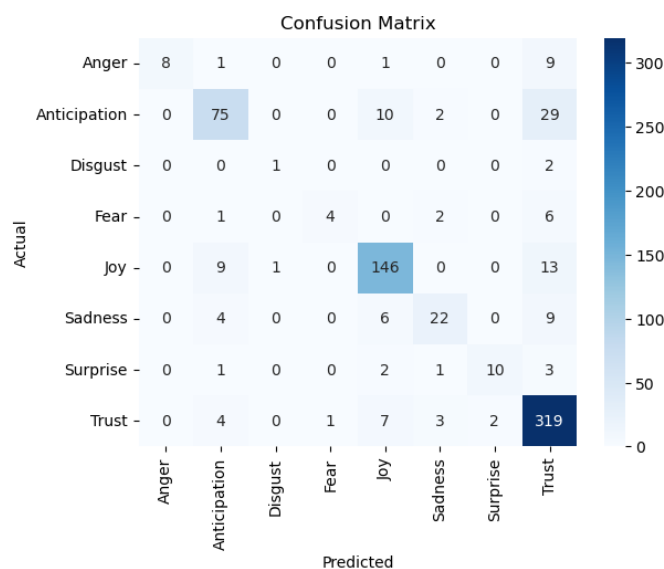
e. **Skenario 5: Non-Sampling, 8 Emosi, PCA**



Gambar 4. 10. *Cumulative explained variance* pada *non-sampling* 8 emosi

*Dataset* tidak dilakukan *resampling* dengan menggunakan 10 emosi. Jumlah fitur awal 4414 dengan PCA menjadi 1500 fitur berdasarkan grafik yang ditunjukkan pada Gambar 4.10. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.11 dapat dihitung hasil evaluasi model, yaitu *accuracy* 81,93%, *precision* 79,03%, *recall*

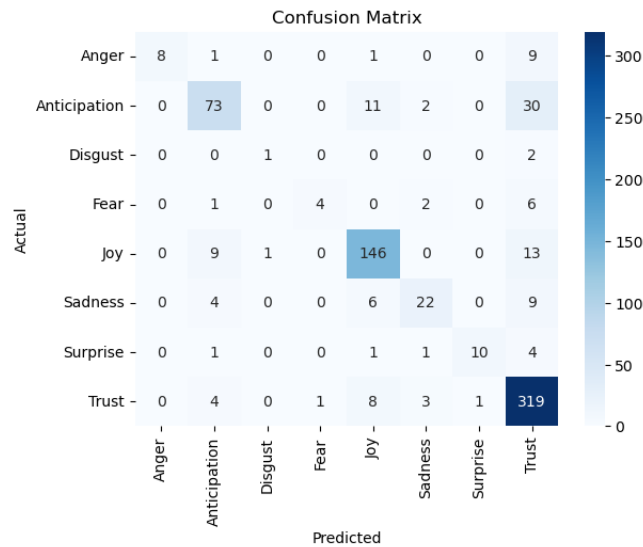
58,08%, dan *F1-score* 64,9%. Artinya, model berhasil 81,93% data dengan benar dari keseluruhan data, dimana hanya 79,03% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



Gambar 4. 11. *Confusion matrix* skenario 5

f. **Skenario 6: *Non-Sampling*, 8 Emosi, Non seleksi fitur dan PCA**

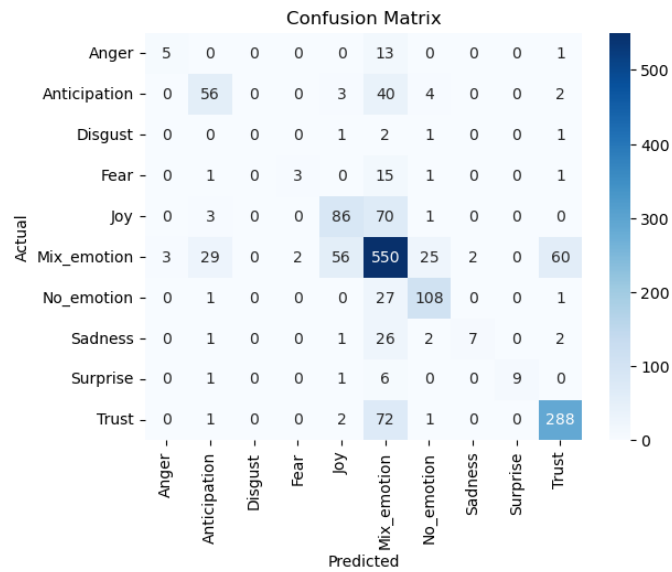
*Dataset* tidak dilakukan *resampling* dengan menggunakan 8 emosi. Jumlah fitur awal 4414 tanpa seleksi fitur dan PCA. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.12 dapat dihitung hasil evaluasi model, yaitu *accuracy* 81,65%, *precision* 79,81%, *recall* 57,86%, dan *F1-score* 64,99%. Artinya, model berhasil 81,65% data dengan benar dari keseluruhan data, dimana hanya 79,81% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



Gambar 4. 12. *Confusion matrix* skenario 6

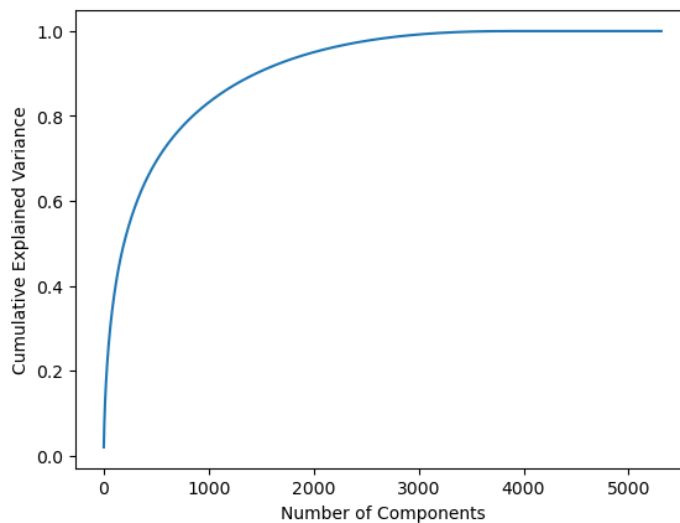
g. **Skenario 7: Sampling, 10 Emosi, Seleksi Fitur**

*Dataset* dilakukan *resampling* dengan menggunakan SMOTE() pada 10 kelas emosi. Jumlah fitur awal 4862 dengan seleksi fitur menjadi 3500 fitur. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.13 dapat dihitung hasil evaluasi model, yaitu *accuracy* 69,76%, *precision* 64,12%, *recall* 45,21%, dan *F1-score* 49,87%. Artinya, model berhasil 69,76% data dengan benar dari keseluruhan data, dimana hanya 64,12 di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



Gambar 4. 13. *Confusion matrix* skenario 7

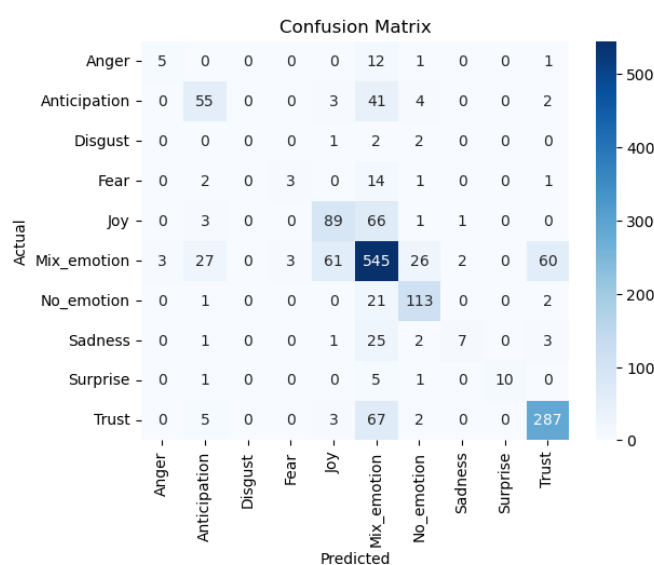
#### h. Skenario 8: *Sampling*, 10 Emosi, PCA



Gambar 4. 14. *Cumulative explained variance* pada *sampling* 10 emosi

*Dataset* dilakukan *resampling* dengan menggunakan SMOTE() pada 10 kelas emosi. Jumlah fitur awal 4862 dengan PCA menjadi 3000 fitur berdasarkan grafik yang ditunjukkan pada Gambar 4.14. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.15 dapat

dihitung hasil evaluasi model, yaitu *accuracy* 69,88%, *precision* 61,94%, *recall* 46,16%, dan *F1-score* 50,2%. Artinya, model berhasil 69,88% data dengan benar dari keseluruhan data, dimana hanya 61,94% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.

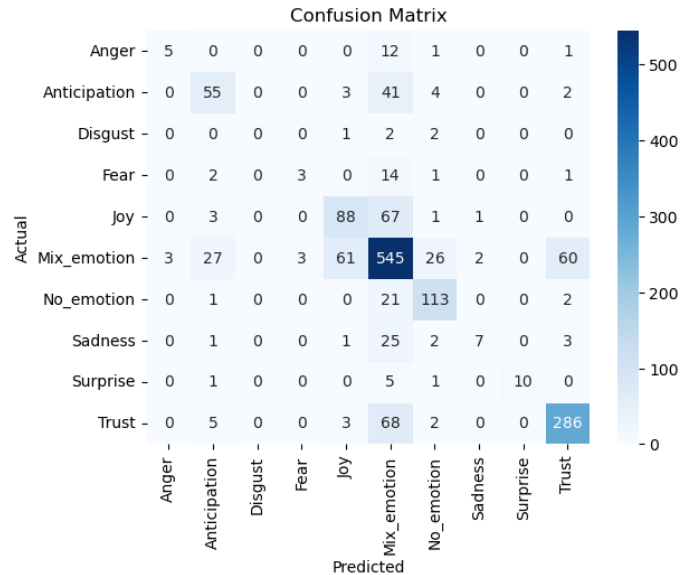


Gambar 4. 15. *Confusion matrix* skenario 8

i. **Skenario 9: *Sampling*, 10 Emosi, Non seleksi fitur dan PCA**

*Dataset* dilakukan *resampling* dengan menggunakan SMOTE() pada 10 kelas emosi. Jumlah fitur awal 4862 tanpa seleksi fitur dan PCA. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.16 dapat dihitung hasil evaluasi model, yaitu *accuracy* 69,76%, *precision* 61,89%, *recall* 46,07%, dan *F1-score* 50,12%. Artinya, model berhasil 69,76% data dengan benar dari keseluruhan data, dimana hanya 61,89% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan

model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.

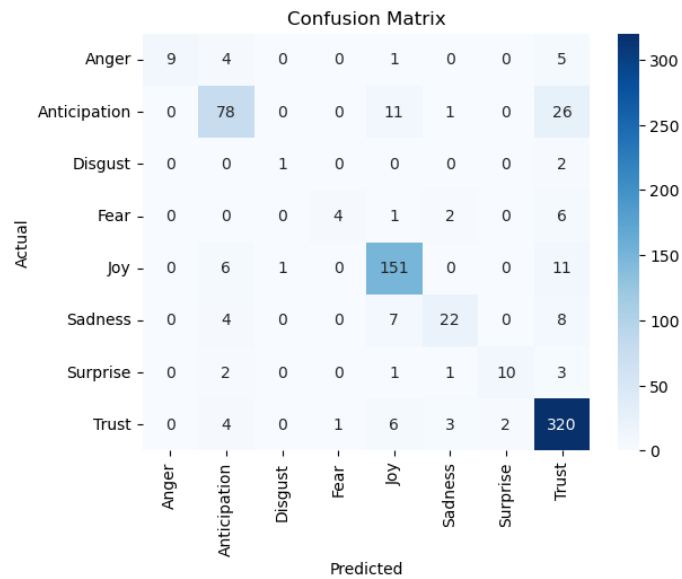


Gambar 4. 16. *Confusion matrix* skenario 9

j. **Skenario 10: *Sampling*, 8 Emosi, Seleksi Fitur**

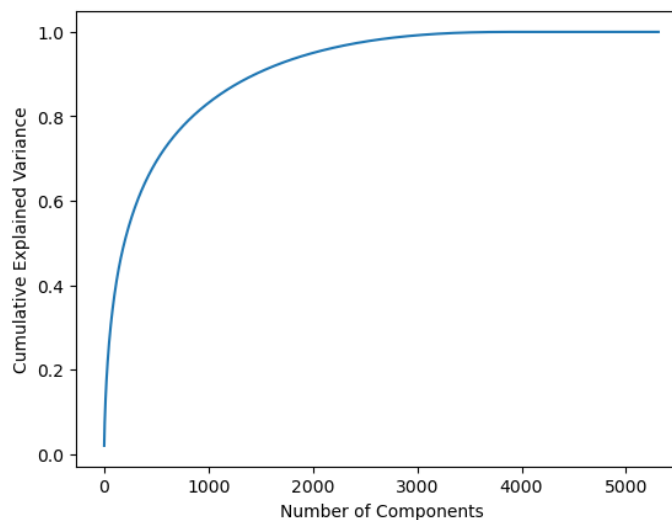
*Dataset* dilakukan *resampling* dengan menggunakan SMOTE() pada 8 kelas emosi. Jumlah fitur awal 3687 dengan seleksi fitur menjadi 2200 fitur. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.17 dapat dihitung hasil evaluasi model, yaitu *accuracy* 83,3%, *precision* 79,7%, *recall* 59,47%, dan *F1-score* 66,21%. Artinya, model berhasil 83,3% data dengan benar dari keseluruhan data, dimana hanya 79,7% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.





Gambar 4. 17. *Confusion matrix* skenario 10

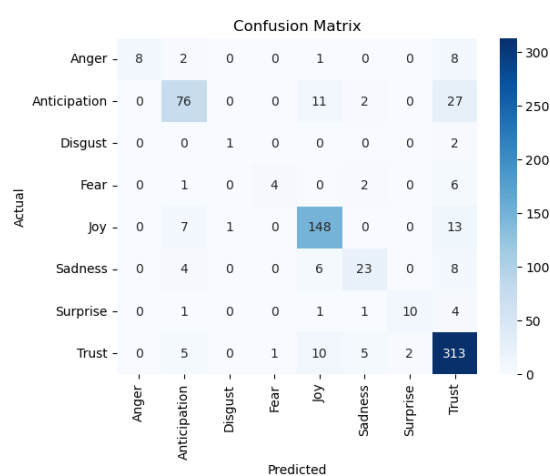
k. **Skenario 11: Sampling, 8 Emosi, PCA**



Gambar 4. 18. *Cumulative explained variance* pada sampling 8 emosi

*Dataset* dilakukan *resampling* dengan menggunakan SMOTE() pada 8 kelas emosi. Jumlah fitur awal 3687 dengan PCA menjadi 1500 fitur berdasarkan grafik yang ditunjukkan pada Gambar 4.18. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.19 dapat

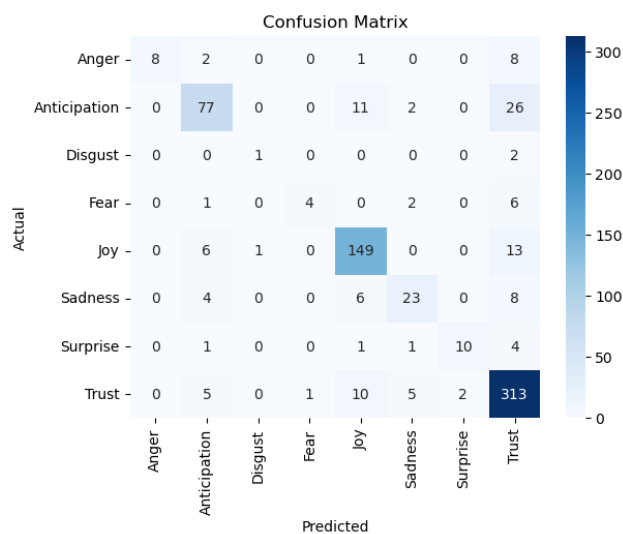
dihitung hasil evaluasi model, yaitu *accuracy* 81,65%, *precision* 78,49%, *recall* 58,42%, dan *F1-score* 64,92%. Artinya, model berhasil 81,65% data dengan benar dari keseluruhan data, dimana hanya 78,49% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



Gambar 4. 19. *Confusion matrix* skenario 11

#### 1. Skenario 12: Sampling, 8 Emosi, Non seleksi fitur dan PCA

*Dataset* dilakukan *resampling* dengan menggunakan SMOTE() pada 8 kelas emosi. Jumlah fitur awal 3687 tanpa seleksi fitur dan PCA. Berdasarkan hasil *confusion matrix* yang ditunjukkan pada Gambar 4.20 dapat dihitung hasil evaluasi model, yaitu *accuracy* 81,93%, *precision* 78,66%, *recall* 58,6%, dan *F1-score* 65,09%. Artinya, model berhasil 81,93% data dengan benar dari keseluruhan data, dimana hanya 78,66% di antaranya yang benar-benar tepat. Nilai *F1-Score* yang cukup rendah dan jauh berbeda dengan akurasi menunjukkan model belum terlalu bagus. *Recall* rendah menunjukkan model tidak mampu menemukan jawaban kelas data berdasarkan pola-pola yang ditemukan dari proses *training*.



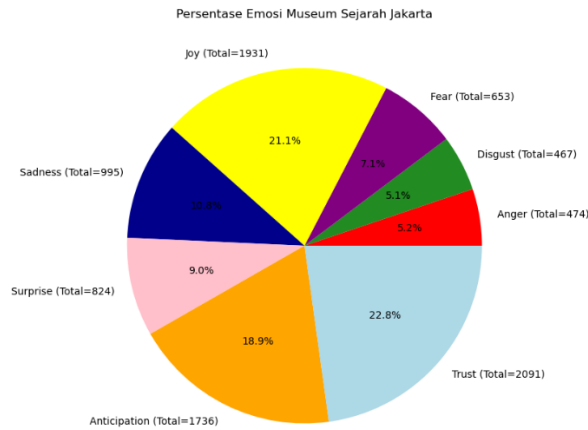
Gambar 4. 20. *Confusion matrix* skenario 12

#### 4.1.5 Analisis Emosi

Setelah dilakukan proses ekstraksi emosi untuk memberi label polaritas emosi, data ulasan yang telah berlabel pada Tabel 4.10 dilakukan perhitungan total frekuensi kemunculan emosi pada setiap objek museum. Hasil perhitungan tersebut dapat divisualisasikan menggunakan grafik dalam bentuk diagram batang dan *pie*. Hasil ekstraksi emosi untuk setiap museum dalam penelitian ini sebagai berikut.

##### 1) *Jakarta History Museum* atau Museum Sejarah Jakarta

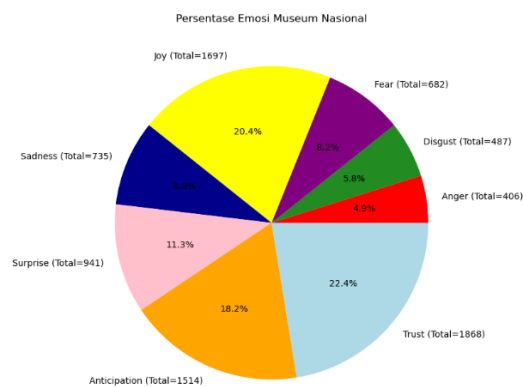
Terdapat 1346 ulasan yang berhasil dikumpulkan dari objek museum ini. Pada Gambar 4.21 menunjukkan bahwa pengunjung museum mendapatkan pengalaman baik di *Jakarta History Museum*, yang tercermin dari persentase kumulatif emosi yakin dan bahagia sebesar 43,9%. Sebanyak 22,8% dari pengunjung merasa yakin dan 21,1% dari pengunjung merasakan bahagia selama kunjungan. Persentase kumulatif tersebut lebih tinggi dibandingkan dengan akumulasi emosi sedih (10,8%), muak atau jijik (5,1%), takut (7,1%), dan marah (5,2%) sebesar 28,2%. Emosi antisipatif (18,9%) dan terkejut (9%) didapat hasil sebesar 27,9%.



Gambar 4. 21. Frekuensi emosi pada *Jakarta History Museum*

## 2) Museum Nasional

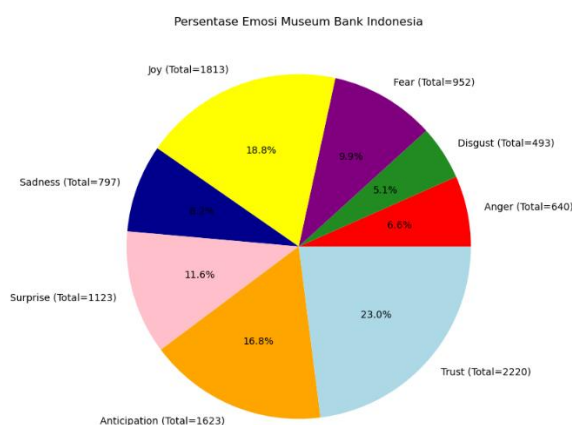
Terdapat 997 ulasan yang berhasil dikumpulkan dari objek museum ini. Pada Gambar 4.22 menunjukkan bahwa pengunjung museum mendapatkan pengalaman baik di Museum Nasional, yang tercermin dari persentase kumulatif emosi yakin dan bahagia sebesar 42,8%. Sebanyak 22,4% dari pengunjung merasa yakin dan 20,4% dari pengunjung merasakan bahagia selama kunjungan. Persentase kumulatif tersebut lebih tinggi dibandingkan dengan akumulasi emosi sedih (8,8%), muak atau jijik (5,8%), takut (8,2%), dan marah (4,9%) sebesar 27,7%. Emosi antisipatif (18,2%) dan terkejut (11,3%) didapat hasil sebesar 29,5%.



Gambar 4. 22. Frekuensi emosi pada Museum Nasional

### 3) Museum Bank Indonesia

Sebanyak 889 ulasan telah berhasil terkumpul untuk objek museum ini. Gambar 4.23 menunjukkan bahwa kunjungan ke Museum Bank Indonesia didominasi pengalaman kunjungan yang baik, hal ini dapat dilihat dari persentase kumulatif emosi yakin dan bahagia sebesar 41,8%. Adapun, 23% dari pengunjung menyatakan perasaan yakin, sementara 18,8% merasakan kebahagiaan selama kunjungan mereka. Persentase kumulatif ini melebihi total akumulasi emosi sedih (8,2%), muak atau jijik (5,1%), takut (9,9%), dan marah (6,6%) yang mencapai 29,8%. Grafik juga menunjukkan bahwa emosi antisipatif (16,8%) dan terkejut (11,6%) berkontribusi sebesar 28,4% dalam pengalaman pengunjung museum.

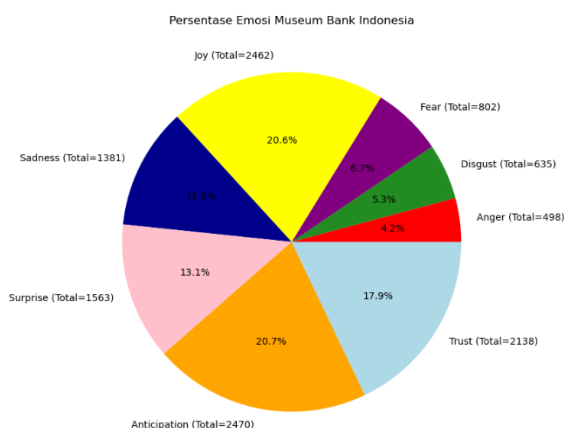


Gambar 4. 23. Frekuensi emosi pada Museum Bank Indonesia

### 4) Museum MACAN

Sebanyak 1131 ulasan telah berhasil terkumpul untuk objek-objek di museum ini. Museum ini didominasi pengalaman baik yang dirasakan pengunjung selama kunjungan. Hal ini dibuktikan dengan persentase kumulatif emosi yakin dan bahagia sebesar 38,5%. Sebesar 20,6% pengunjung merasa bahagia dan 17,9% pengunjung merasa yakin. Persentase kumulatif tersebut melebihi total akumulasi emosi sedih

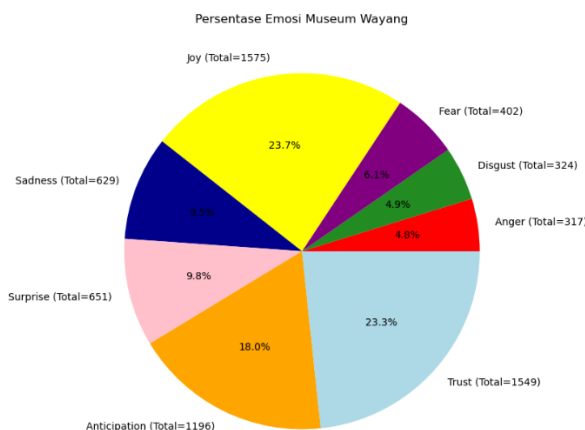
(11,6%), muak atau jijik (5,3%), takut (6,7%), dan marah (4,2%) sebesar 27,8%. Emosi antisipatif (20,7%) dan terkejut (13,1%) berkontribusi sebesar 33,8% dalam pengalaman pengunjung museum. Perbandingan persentase frekuensi setiap emosi pada Museum MACAN ditunjukkan pada Gambar 4.24.



Gambar 4. 24. Persentase frekuensi emosi pada Museum MACAN

#### 5) Museum Wayang

Sebanyak 974 ulasan telah berhasil terkumpul untuk objek-objek di museum ini. Museum ini didominasi pengalaman baik yang dirasakan pengunjung selama kunjungan. Hal ini dibuktikan dengan persentase kumulatif emosi yakin dan bahagia sebesar 47%. Emosi bahagia memiliki persentase tertinggi sebesar 23,7% diikuti emosi yakin sebesar 23,3%. Persentase kumulatif tersebut melebihi total akumulasi emosi sedih (9,5%), muak atau jijik (4,9%), takut (6,1%), dan marah (4,8%) sebesar 25,3%. Emosi antisipatif (18%) dan terkejut (9,8%) berkontribusi sebesar 27,8% dalam pengalaman pengunjung museum. Perbandingan persentase frekuensi setiap emosi pada Museum Wayang ditunjukkan pada Gambar 4.25.



Gambar 4. 25. Persentase frekuensi emosi pada Museum Wayang

#### 4.1.6 Analisis Persepsi

Analisis persepsi pengunjung terhadap pengalaman ataupun museumnya itu sendiri menggunakan analisis *word cloud*. Proses visualisasi *word cloud* ditunjukkan pada Lampiran 15. Visualisasi *word cloud* dilakukan menggunakan *library wordcloud*, dimana hanya mengambil kata benda dan kata sifat menggunakan *pos\_tag*. Hasil analisis persepsi setiap objek museum sebagai berikut.

##### 1) *Jakarta History Museum* atau Museum Sejarah Jakarta

Gambar 4.26 menunjukkan hasil visualisasi data ulasan pengunjung *Jakarta History Museum* menggunakan *word cloud*. Berdasarkan gambar tersebut, dapat diketahui bahwa kata “sejarah” dan “bagus” merupakan kata yang paling dominan. Artinya, *Jakarta History Museum* merupakan salah satu objek wisata bersejarah yang bagus, dimana museum ini menampilkan sejarah perkembangan, baik kejadian maupun masyarakat Kota Jakarta. Pada hasil visualisasi tersebut juga terdapat kata “bersih”, “ramai”, “bangunan”, dan “foto”, hal ini dapat menjelaskan bahwa *Jakarta History Museum* menjadi museum yang bersih dan ramai pengunjung yang tertarik mengeksplorasi sejarah. *Jakarta History Museum* merupakan salah satu bangunan ikonik di Kota Tua sehingga banyak pengunjung datang untuk berfoto.



Gambar 4. 26. *Wordcloud* Jakarta History Museum

## 2) Museum Nasional

Hasil visualisasi menggunakan *word cloud* Museum Nasional dapat dilihat pada Gambar 4.27. Berdasarkan gambar tersebut, terlihat bahwa kata "bagus", "sejarah", dan "Indonesia" yang menggambarkan museum ini memiliki koleksi peninggalan mengenai sejarah Indonesia. Selain itu, juga terdapat kata "bersih" dan "nyaman" yang menjelaskan kondisi dan suasana museum. Kata "anak" dapat menjelaskan bahwa museum ini cocok untuk anak-anak belajar sejarah Indonesia dengan melihat koleksi-koleksi museum. Kata "koleksi", "artefak", "patung", dan "arca" dapat menggambarkan museum ini memiliki koleksi lengkap berupa artefak, patung, atau arca yang merupakan peninggalan sejarah Indonesia. Kemudian, kata "pameran" dan "imersifa" menggambarkan bahwa museum nasional sering mengadakan pameran, misalnya pameran kopi togetherness dan corat-coret asyik. Serta, pada



museum ini juga memiliki ruang imersifa dengan proyeksi *video mapping* 360° yang menampilkan alam, seni, dan budaya Indonesia.



Gambar 4. 27. Wordcloud Museum Nasional

### 3) Museum Bank Indonesia

Hasil visualisasi menggunakan *word cloud* Museum Bank Indonesia dapat dilihat pada Gambar 4.28. Kata "sejarah" dan "uang" terlihat dominan. Museum Bank Indonesia sendiri merupakan museum yang memiliki banyak koleksi, seperti halnya koleksi mata uang kerajaan di Indonesia, uang kolonial, uang awal kemerdekaan, uang token, uang khusus, dan sebagainya. Kemudian, kata "bagus", "nyaman", "bersih", "menarik" dapat menjelaskan kondisi dan suasana di museum.



Gambar 4. 28. Wordcloud Museum Bank Indonesia

#### 4) Museum MACAN

Hasil visualisasi menggunakan *word cloud* Museum MACAN dapat dilihat pada Gambar 4.29. Berdasarkan hasil *wordcloud* tersebut, kata "foto", "bagus", "seni", "pameran" terlihat paling dominan. Dari kata-kata tersebut dapat menjelaskan bahwa Museum MACAN merupakan salah satu museum yang memiliki koleksi tentang karya seni modern. Banyak pengunjung datang ke museum ini untuk melihat dan berfoto pada pameran karya seni tersebut. Dimana kata "instagramable" juga terlihat di hasil *wordcloud*, hal ini menunjukkan museum MACAN memiliki koleksi seni yang bagus sehingga banyak spot foto yang *instagramable*.



Gambar 4. 29. *Wordcloud* Museum Macan

#### 5) Museum Wayang

Hasil visualisasi menggunakan *word cloud* Museum Wayang dapat dilihat pada Gambar 4.30. Berdasarkan gambar tersebut dapat diketahui bahwa kata "koleksi", "boneka" dan "bagus" merupakan kata yang dominan. Museum wayang sendiri memiliki koleksi-koleksi wayang atau boneka dari berbagai daerah di Indonesia. Serta, juga terdapat koleksi wayang dari luar negeri. Selain itu, dari Gambar 4.30 juga menunjukkan kata "anak", artinya museum ini cocok bagi anak-anak untuk melihat dan mempelajari sejarah wayang. Kata "menarik" dan "nyaman" dapat menjelaskan suasana museum. Serta, terdapat kata



Museum	Emosi (%)							
	<i>Trust</i>	<i>Joy</i>	<i>Surprise</i>	<i>Anticipation</i>	<i>Anger</i>	<i>Sad</i>	<i>Disgust</i>	<i>Fear</i>
Museum MACAN	17,9	20,6	13,1	20,7	4,2	11,6	5,3	6,7
Museum Wayang	23,3	23,7	9,8	18	4,8	9,5	4,9	6,1
Rata-rata	21,88	20,92	10,96	18,52	5,14	9,78	5,24	7,6

Dari Tabel 4.12 dapat disimpulkan bahwa Museum Wayang mendapatkan nilai emosi *trust* dan *joy* tertinggi dibandingkan museum lainnya dengan persentase masing-masing emosi sebesar 23,3% dan 23,7%. Emosi *anticipation* dan *surprise* tertinggi dimiliki Museum MACAN dengan masing-masing persentase sebesar 13,1% dan 20,7%. Selanjutnya, Museum Bank Indonesia mendapatkan persentase emosi *anger* tertinggi, yaitu sebesar 6,6%. Kemudian, Museum MACAN juga menjadi museum yang mendapatkan persentase emosi *sad* tertinggi sebesar 11,6%. Emosi *disgust* banyak diberikan pengunjung Museum Nasional dengan persentase sebesar 5,8%. Terakhir, Museum Bank Indonesia sebagai museum yang memiliki persentase emosi *fear* tertinggi, yaitu sebesar 9,9%.

Berdasarkan hasil tersebut juga menunjukkan lima objek museum membuat pengunjung merasakan positif (*joy* dan *trust*) dengan rentang persentase antara 17% hingga 23%. Emosi *anticipation* antara 16% hingga 20% dan *surprise* antara 9% dan 13%. Sedangkan, emosi negatif (*sadness*, *fear*, *disgust*, dan *anger*) terhadap lima objek museum rendah, di bawah 12%. Hal ini dapat menggambarkan bahwa pengunjung mendapatkan emosi positif yang lebih tinggi selama melakukan kunjungan di setiap museum. Rata-rata persentase emosi *trust* (21,88%), *joy* (20,92%), *anticipation* (18,52%), dan *surprise* (10,96%) lebih tinggi dibandingkan dengan persentase emosi *sadness* (9,78%), *fear* (7,6%), *disgust* (5,24%), dan *anger* (5,14%).

#### 4.2.2 Pembahasan Hasil Analisis Persepsi

Tabel 4.13 menunjukkan perbandingan persepsi dari setiap objek museum. Tabel tersebut menunjukkan kata kunci dan deskripsi yang menggambarkan museum berdasarkan ulasan yang diberikan pengunjung. Berdasarkan tabel tersebut dapat diketahui bahwa pengunjung memiliki pandangan terhadap masing-masing museum, dimana setiap museum memiliki keunikan dan daya tariknya sendiri, mulai dari aspek sejarah, kebersihan, hingga koleksi seni dan pertunjukan yang disajikan.

Tabel 4. 13. Perbandingan persepsi setiap museum

<b>Museum</b>	<b>Kata kunci</b>	<b>Hasil</b>
<i>Jakarta History Museum</i>	sejarah, bagus, bersih, ramai, bangunan, dan foto	<i>Jakarta History Museum</i> fokus pada pameran dan informasi seputar perkembangan sejarah, termasuk kejadian dan kehidupan masyarakat Kota Jakarta. Museum ini menjadi museum yang bersih dan ramai pengunjung yang tertarik mengeksplorasi sejarah dengan bangunan ikonik sebagai daya tarik visual museum untuk pengunjung berfoto.
Museum Nasional	bagus, sejarah, Indonesia, bersih, nyaman, anak, koleksi, pameran, dan imersifa	Museum Nasional memiliki fokus pada sejarah Indonesia dengan keberagaman koleksi museum, seperti artefak, patung, dan arca. Museum Nasional cocok untuk anak-anak. Dimana museum ini aktif dalam menyelenggarakan pameran serta menyediakan pengalaman interaktif melalui ruang imersifa.

Museum	Kata kunci	Hasil
Museum Bank Indonesia	sejarah, uang, bagus, nyaman, bersih, dan menarik	Museum Bank Indonesia memiliki penekanan utama pada aspek sejarah uang di Indonesia. Dengan koleksi yang kaya terkait dengan mata uang dari berbagai periode sejarah. Museum ini menjaga kebersihan, kenyamanan, dan daya tarik pengunjung.
Museum MACAN	foto, seni, bagus, pameran, dan <i>instagramable</i>	Museum MACAN memiliki fokus pada seni modern yang menampilkan koleksi dan pameran seni kontemporer. Museum MACAN berhasil menarik pengunjung yang menginginkan pengalaman berfoto yang unik dan menarik untuk kepentingan konten media sosial.
Museum Wayang	koleksi, boneka, bagus, anak, menarik, nyaman, pewayangan, dan pertunjukan	Museum Wayang memiliki fokus pada koleksi-koleksi wayang dan boneka. Museum ini cocok sebagai destinasi edukasi bagi anak-anak untuk melihat dan mempelajari sejarah wayang. Museum Wayang berusaha memberikan pengalaman yang tidak hanya edukatif, tetapi juga menyenangkan dan nyaman dengan menyediakan pertunjukan langsung dan aktivitas pewayangan.

#### 4.2.3 Pembahasan Hasil Klasifikasi dengan SVM

Hasil klasifikasi emosi pada data ulasan pengunjung museum menggunakan SVM dapat dilihat pada Tabel 4.14 sedangkan hasil klasifikasi emosi pada data

ulasan yang telah dilakukan *sampling* dengan SMOTE ditunjukkan pada Tabel 4.15. Model SVM menggunakan 8 emosi dan data yang telah di-*sampling* menggunakan SMOTE dengan dilakukan seleksi fitur merupakan model dengan performa yang paling baik dibandingkan model lainnya dengan metrik evaluasi, yaitu *accuracy* 83,3%, *precision* 79,7%, *recall* 59,47%, dan *F1-Score* 66,21%. Artinya, model berhasil memprediksi 83,3% data dengan benar dari keseluruhan data, dimana hanya 79,7% di antaranya yang benar-benar tepat.

Tabel 4. 14. Perbandingan hasil klasifikasi sebelum *sampling*

Metrik Evaluasi	10 Emosi			8 Emosi		
	Seleksi Fitur	Non-Seleksi Fitur	PCA	Seleksi Fitur	Non-Seleksi Fitur	PCA
<i>Accuracy</i>	70,32%	69,69%	69,69%	83,05%	81,65%	81,93%
<i>Precision</i>	65,9%,	64,15%	64,15%	74,8%	79,81%	79,03%
<i>Recall</i>	42,5%,	44,35%	44,35%	53,59%	57,86%	58,08%
<i>F1-Score</i>	48,26%.	49,19%	49,19%	60,25%	64,99%	64,9%

Tabel 4. 15. Perbandingan hasil klasifikasi setelah *sampling*

Metrik Evaluasi	10 Emosi			8 Emosi		
	Seleksi Fitur	Non-Seleksi Fitur	PCA	Seleksi Fitur	Non-Seleksi Fitur	PCA
<i>Accuracy</i>	69,76%	69,76%	69,88%	83,3%	81,93%	81,65%
<i>Precision</i>	64,12%	61,89%	61,94%	79,7%	78,66%	78,49%
<i>Recall</i>	45,21%	46,07%	46,16%	59,47%	58,6%	58,42%
<i>F1-Score</i>	49,87%	50,12%	50,2%	66,21%	65,09%	64,92%

Berdasarkan Tabel 4.14 dan 4.15 dapat diketahui bahwa metode seleksi fitur cenderung memberikan performa yang lebih baik dibandingkan dengan non-seleksi fitur dalam hal akurasi dan presisi. Pada skenario *sampling* 8 emosi, PCA memberikan performa yang sedikit lebih rendah dibandingkan dengan seleksi fitur dalam hal akurasi, presisi, *recall*, dan *F1-score*. Secara umum, skenario 8 emosi memiliki performa yang lebih baik dibandingkan dengan skenario 10 emosi. Terlihat dari peningkatan nilai akurasi, presisi, *recall*, dan *F1-score* pada sebagian besar metode evaluasi.