

## BAB II

### TINJAUAN PUSTAKA DAN LANDASAN TEORI

#### 2.1 Tinjauan Pustaka

Penelitian-penelitian terdahulu yang berkaitan dan memiliki relevansi terhadap penelitian yang dilakukan akan ditinjau sebagai bahan referensi dan perbandingan untuk menyempurnakan konsep dan kerangka penelitian. Tabel 2.1 menunjukkan hasil tinjauan terhadap penelitian-penelitian terkait.

Penelitian yang dilakukan oleh Samer Muthana Sarsam, dkk., pada tahun 2021 terkait ekstraksi dan analisis karakteristik emosi pada konten bunuh diri menggunakan metode *NRC Affect Intensity Lexicon* dan *SentiStrength technique* [23]. Data penelitian ini diambil dari Twitter dengan jumlah total 4.987 *tweet* dalam bahasa Inggris. Hasil penelitian ini menunjukkan *tweets* terkait dengan konten bunuh diri mengandung emosi ketakutan (*fear*), kesedihan (*sad*), dan sentimen negatif.

Penelitian yang dilakukan oleh Agus Sasmito Aribowo dan Siti Khomsah pada tahun 2021 terkait penerapan *text mining* untuk mengidentifikasi emosi pengguna [24]. Penelitian ini bertujuan untuk mengetahui jenis emosi apa saja yang diekspresikan pengguna Twitter di Indonesia selama pandemi Covid-19. Serta, untuk mengetahui bahan percakapan atau kata yang muncul terus-menerus untuk setiap jenis emosi. Penelitian ini menggunakan pendekatan berbasis *lexicon* untuk mengekstrak emosi. Data penelitian ini menggunakan *tweet* dalam Bahasa Indonesia dari mulai April hingga Agustus 2020. Hasil dari penelitian ini berupa grafik perkembangan kondisi emosi masyarakat terkait pandemi Covid-19 untuk setiap bulannya dalam *dataset*.

Penelitian yang dilakukan oleh Siti Khomsah, dkk. pada tahun 2021 terkait penerapan *big data analytics* untuk analisis sentimen, emosi, dan persepsi wisatawan [25]. Penelitian ini bertujuan untuk mengetahui sentimen, emosi, dan persepsi dari wisatawan terhadap destinasi wisata di Purwokerto. Data penelitian ini diambil melalui *web scraping* Google Maps Reviews. Hasil dari penelitian ini

diketahui 5 dari 8 destinasi wisata memiliki sentiment positif yang tinggi (>70%). Emosi yang paling dominan adalah *joy* dan *trust*. Serta, visualisasi *word cloud* untuk setiap destinasi wisata.

Penelitian yang dilakukan oleh Faheem Aslam, dkk. pada tahun 2020 terkait analisis sentimen dan emosi yang ditimbulkan oleh judul berita tentang wabah penyakit Covid-19 [26]. Penelitian ini bertujuan untuk mengesktrak dan mengklasifikasikan sentimen dan emosi dari judul berita berbahasa Inggris terkait penyakit Covid-19. Data penelitian ini menggunakan 141.208 judul berita berbahasa Inggris. Hasil penelitian ini menunjukkan judul berita memiliki skor emosi negatif yang tinggi. Dari seluruh judul berita, 52% mengandung sentiment negatif, 30% mengandung sentiment positif, dan 18% bersifat netral. Emosi dominan yang terkandung, yakni takut, yakin, antisipatif, sedih, dan marah.

Penelitian yang dilakukan oleh Dhafar Hamed Abd, dkk., pada tahun 2021 mengenai sistem analisis sentimen pandangan dan persepsi orang menggunakan pendekatan berbasis *lexicon* pada ulasan film [27]. Data penelitian menggunakan 50.000 ulasan film di IMDb. Penelitian ini menghasilkan sentimen *lexicon* sebanyak 25.000, 30.000, 34.000, dan 40.000 istilah ulasan. Serta, penelitian mencapai hasil akurasi prediksi terbaik sebesar 76,585% dalam menganalisis sentiment.

Penelitian yang dilakukan oleh Muhammad Afzaal, dkk., pada tahun 2018 mengenai mengklasifikasi pendapat berbasis multiaspek untuk ulasan wisatawan yang bertujuan untuk mengekstrak aspek implisit eksplisit serta mengklasifikasikan pendapat multi aspek [28]. Penelitian ini menggunakan data dari berbagai sumber yang berbeda, yaitu SemEval ABSA sebanyak 425 *reviews*, We8There sebanyak 4.488 *reviews*, TripAdvisor sebanyak 44.000 *reviews*, Parks sebanyak 3.000 *reviews*, dan Art Museums sebanyak 5.000 *reviews*. Penelitian ini menggunakan pendekatan *lexicon* untuk ekstraksi informasi dari ulasan dan menggunakan *binary relevance*, *classifier chains*, *multilabel k-NN*, *RAkEL*, dan *ensemble classifier chains* untuk klasifikasi. Hasil penelitian ini menunjukkan akurasi terbaik sebesar 90% per label pada klasifikasi ketika mengekstraksi 87% aspek relevan.

Penelitian yang dilakukan oleh Arif Nur Rohman, dkk. pada tahun 2019 berkaitan dengan deteksi emosi pada komentar dan postingan di Facebook menggunakan pendekatan *lexicon* dan NLP [29]. Penelitian ini bertujuan untuk mengidentifikasi emosi yang terkandung pada komentar dan postingan. *Dataset* penelitian ini diambil melalui *scraping timeline* pengguna Facebook berjumlah 100 teks postingan dalam bahasa Indonesia. Penelitian ini menggunakan *NRC Emolex* disertai perluasan *lexicon Emolex* dengan Kateglo. Hasil penelitian ini menunjukkan 61,53% label emosi yang akurat dengan perluasan kosakata sebanyak 34.872 kata.

Penelitian yang dilakukan oleh Amir Hamzah pada tahun 2021 berkaitan dengan deteksi emosi menggunakan pendekatan *lexicon* untuk hasil kuesioner akademik [30]. Penelitian ini bertujuan untuk implementasi metode *NRC Emolex* untuk mendeteksi emosi pada teks komentar mahasiswa hasil kuesioner. Penelitian ini menggunakan *dataset* berasal dari kuesioner evaluasi pembelajaran sebanyak 3.975 komentar. Penelitian ini menunjukkan hasil, 46,7% dari seluruh data dapat diketahui labelnya. Emosi sedih, bahagia, dan takut merupakan top tiga emosi dengan persentase tertinggi.

Penelitian yang dilakukan oleh Helen Sastypratiwi, dkk. pada tahun 2022 berkaitan dengan penerapan algoritma SVM dengan *tuning hyperparameter* PSO untuk klasifikasi emosi pada lirik lagu [31]. Penelitian ini bertujuan untuk menentukan kelas emosi yang terkandung di lirik lagu berbahasa Indonesia. Penelitian ini menggunakan lirik lagu sebagai *dataset* dimana menjadi 3 *dataset*, yakni per bait, per baris, dan keseluruhan lagu. Hasil penelitian ini menunjukkan skenario SVM+PSO dengan *dataset* per baris memiliki performa model yang paling baik dengan akurasi mencapai 92,13%.

Penelitian yang dilakukan oleh Rahmadya Trias Handayanto, dkk. pada tahun 2021 berkaitan dengan analisis sentimen ulasan pelanggan Summarecon Mal Bekasi pada Google Review [21]. Penelitian ini bertujuan untuk mengetahui kepuasan pelanggan dengan menganalisis sentimen dari ulasan-ulasan pelanggan Summarecon Mal Bekasi. *Dataset* berjumlah 2.143 komentar diambil dengan *scraping* dan dilakukan pelabelan dengan kamus *lexicon*. Penelitian ini

menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM). Hasil penelitian ini adalah akurasi metode SVM lebih baik dibandingkan *Naïve Bayes*, yakni SVM mencapai akurasi 100% sedangkan *Naïve Bayes* hanya mencapai 80.95%.

Penelitian yang dilakukan oleh Toni Arifin Fitriyani pada tahun 2020 terkait penggunaan *N-gram* untuk analisis sentiment menggunakan metode SVM pada aplikasi Sambara [22]. Penelitian ini bertujuan untuk mendeteksi polaritas pada ulasan berupa komentar negatif atau positif dengan menggunakan *text mining*. Penelitian ini menggunakan metode SVM untuk klasifikasi sentimen. Hasil penelitian ini adalah akurasi tertinggi dengan menggunakan tri-gram dengan jumlah *dataset* 1200 sebesar 92% dengan nilai AUC 0.954.

Penelitian yang dilakukan oleh Fransiska Vina Sari dan Arief Wibowo pada tahun 2019 berkaitan dengan analisis sentimen opini pelanggan JD.ID [32]. Penelitian ini bertujuan untuk mengoptimalkan model dan mengetahui analisis sentiment opini yang diberikan konsumen tentang JD.ID. Penelitian ini menggunakan pembobotan TF-IDF yang disertai penambahan fitur konversi emotikon emosi dan algoritma *Naïve Bayes*. Penelitian menghasilkan akurasi sebesar 96,44% dalam mengklasifikasi sentiment opini pelanggan hanya menggunakan algoritma *Naïve Bayes* saja. Namun, dengan menambahkan pembobotan fitur TF-IDF yang disertai konversi ikon emosi, akurasi meningkat menjadi sebesar 98%.

Berdasarkan tinjauan pustaka, penelitian yang dilakukan memiliki perbedaan yang menonjol dengan penelitian terdahulu. *Dataset* yang digunakan pada penelitian ini menggunakan data ulasan dari Google Maps yang juga digunakan pada beberapa penelitian terdahulu. Ulasan museum sebagai subjek penelitian yang akan membedakan penelitian ini dengan penelitian sebelumnya. Penelitian ini akan menggunakan metode *NRC Emolex* untuk pelabelan *dataset*, SVM untuk proses klasifikasi emosi, serta visualisasi *word cloud* untuk analisis persepsi. Perbedaan penelitian ini dengan terdahulu juga terdapat pada tahapan *preprocessing* dimana penelitian ini akan disertai penambahan konversi emotikon.

Tabel 2. 1. Tinjauan pustaka

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
1.	<i>A Lexicon-based Approach to Detecting Suicide-related messages on Twitter</i>	Menggunakan pendekatan <i>lexicon</i> untuk ekstraksi emosi.	Menggunakan data yang diambil dari Twitter dan topik yang dibahas spesifik mengenai <i>tweet</i> bunuh diri.	Penelitian ini hanya berfokus terhadap <i>tweet</i> berbahasa Inggris.	Penelitian ini menggunakan pendekatan <i>lexicon</i> tepatnya <i>NRC Affect Instensity Lexicon</i> dan <i>SentiStrength</i> agar dapat menganalisis dan ekstraksi emosi pengguna Twitter.	Penelitian ini menunjukkan <i>tweet</i> bunuh diri memiliki emosi takut, sedih, dan sentimen negatif dibandingkan <i>tweet</i> non-bunuh diri.
2.	<i>Implementation of Text Mining for Emotion Detection Using the Lexicon Method (Case Study: Tweets About Covid-19)</i>	Penelitian ini menganalisis emosi dari <i>text</i> menggunakan metode <i>lexicon</i> .	Penelitian ini menggunakan data yang bersumber dari Twitter dan berfokus terhadap cuitan mengenai Covid-19.	Penelitian ini berfokus ekstraksi emosi dari teks dan menghapus <i>symbol emoticon</i> .	Penelitian ini menggunakan metode <i>lexicon</i> untuk mendeteksi emosi dari cuitan pengguna di Twitter terhadap Covid-19.	Penelitian ini menunjukkan grafik perkembangan emosi publik terkait Covid-19. Dimana emosi yang paling dominan adalah emosi takut.
3.	<i>Big Data Analytics to Analyze Sentiment, Emotions, and Perceptions of Travelers (Case Study: Tourism Destination in</i>	Menggunakan <i>NRC Emolex</i> untuk ekstraksi emosi dan <i>word cloud</i> untuk analisis persepsi.	Objek penelitian ini destinasi wisata di Purwokerto.	Penelitian ini sudah mencakup analisis sentimen, emosi, dan persepsi wisatawan. Akan tetapi, belum adanya model klasifikasi.	Penelitian ini menggunakan <i>NRC Emolex</i> untuk menganalisis sentiment dan emosi. Serta, menggunakan <i>word cloud</i>	Hasil penelitian ini menunjukkan secara umum wisatawan memiliki sentimen yang positif. Hanya sedikit wisatawan yang memberikan

No.	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
	<i>Purwokerto (Indonesia)</i>				untuk analisis persepsi.	sentiment negatif. Emosi yang paling dominan adalah bahagia dan yakin.
4.	<i>Sentiments and Emotions Evoked by News Headlines of Coronavirus Disease (COVID-19) outbreak</i>	Menggunakan <i>NRC Emolex</i> untuk ekstraksi emosi.	Pada penelitian ini menggunakan <i>dataset</i> judul berita Covid-19.	Penelitian ini hanya berfokus terhadap judul berita berbahasa Inggris. Serta, tidak memperhatikan emotikon.	Penelitian ini menggunakan <i>NRC Emolex</i> untuk menganalisis sentiment dan emosi.	Penelitian ini menunjukkan judul berita memiliki skor emosi negatif yang tinggi. Dimana 52% judul berita mengandung sentiment negatif, 30% mengandung sentiment positif, dan 18% bersifat netral. Emosi dominan yang terkandung di antaranya takut, yakin, antisipatif, sedih, dan marah.
5.	<i>Analyzing Sentiment System to Specify Polarity by Lexicon-based</i>	Menggunakan pendekatan <i>lexicon</i> untuk ekstraksi emosi.	Menggunakan data <i>movie reviews</i> yang diambil dari IMDb.	Penelitian ini hanya berfokus pada tiga kelas emosi, yakni positif, negatif, dan alami.	Penelitian ini menggunakan <i>lexicon</i> lebih dari 50.000 ulasan film untuk menganalisis sentiment atau opini. Serta,	Penelitian ini menunjukkan hasil akurasi prediksi yang baik dalam menganalisis sentiment terhadap ulasan film.

No.	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
					deteksi polaritas dengan analisis sentimen.	
6.	<i>Multiaspect-based Opinion Classification Model for Tourist Reviews</i>	Menggunakan pendekatan <i>lexicon</i> untuk ekstraksi informasi dari <i>reviews</i> .	Penelitian ini mencakup <i>tourist reviews</i> terhadap berbagai macam tempat wisata, seperti hotel, restoran, museum, dan taman. <i>Dataset</i> yang digunakan diambil dari beberapa platform <i>online</i> , yakni Facebook, TripAdvisor, Zomato, dan Expedia.	Pada penelitian ini membahas <i>tourist reviews</i> dari berbagai macam jenis tempat wisata sehingga memerlukan <i>computational cost</i> yang tidak sedikit dari sisi waktu pemrosesan	Penelitian ini menggunakan <i>lexicon</i> yang bertujuan untuk mengekstraksi informasi dari <i>tourist reviews</i> serta menggunakan algoritma <i>machine learning (ML) multilabel classification</i> untuk klasifikasinya.	Penelitian ini mengesktrak aspek dari <i>reviews</i> atau opini dan mengklasifikasi menjadi beberapa polaritas kelas yang berbeda menggunakan beberapa algoritma ML klasifikasi.
7.	Deteksi Emosi Media Sosial Menggunakan Pendekatan <i>Lexicon</i> dan <i>Natural Language Processing</i>	Mengekstrak emosi menggunakan <i>NRC Emolex</i> .	Pada penelitian ini dilakukan perluasan kosakata menggunakan Kateglo sehingga terdapat 34.872 kosakata. Serta, pada penelitian ini tidak melakukan	Pada penelitian ini membuang <i>emoticon</i> karena hanya membahas teks yang akan diuji pada <i>lexicon</i> . Serta, pada penelitian ini tidak dilakukan <i>stopword removal</i> .	Penelitian ini menggunakan <i>NRC Emolex</i> untuk mengekstraksi emosi disertai dengan perluasan kosakata menggunakan Kateglo.	Penelitian ini menunjukkan 61,53% label emosi yang akurat dengan perluasan kosakata sebanyak 34.872 kata.

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
			pemodelan dengan ML.			
8.	<i>Lexicon-based Emotion Detection for Academic Questionnaire Results</i>	Mengekstrak emosi menggunakan <i>NRC Emolex</i> .	Pada penelitian ini tidak melakukan pemodelan dengan ML. Serta, data penelitian bersumber dari kuesioner evaluasi pembelajaran mahasiswa.	Pada penelitian ini tidak memperhatikan <i>emoticon</i> .	Penelitian ini menggunakan <i>NRC Emolex</i> untuk mendeteksi emosi pada teks komentar hasil kuesioner.	Penelitian ini menunjukkan hasil, 46,7% dari seluruh data dapat diketahui labelnya. Dimana emosi <i>sadness, joy, dan fear</i> merupakan top 3 persentase tertinggi.
9.	Klasifikasi Emosi pada Lirik Lagu Menggunakan Algoritma <i>Multiclass SVM</i> dengan <i>Tuning Hyperparameter PSO</i>	Menggunakan algoritma SVM untuk klasifikasi emosi.	Pada penelitian ini pelabelan <i>dataset</i> dilakukan secara manual oleh 3 orang. Kelas emosi pada penelitian ini, yaitu cinta, senang, marah, takut, dan sedih. Serta, menggunakan <i>tuning hyperparameter PSO</i> .	Pada penelitian ini melakukan pelabelan manual. Jika <i>dataset</i> yang digunakan berjumlah besar maka pelabelan manual membutuhkan tenaga, waktu, dan sumber daya yang lebih besar.	Penelitian ini menggunakan algoritma SVM untuk klasifikasi emosi dari lirik lagu dan PSO untuk <i>tuning hyperparameter</i> .	Penelitian ini berhasil melakukan klasifikasi emosi dari teks lirik lagu berbahasa Indonesia dengan akurasi tertinggi sebesar 92,13%.
10.	Analisis Sentimen Pada Situs Google Review dengan <i>Naïve Bayes</i> dan	Menggunakan metode <i>lexicon</i> untuk pelabelan dan SVM untuk klasifikasi.	Penelitian ini berfokus pada analisis sentimen, yakni kelas positif dan negatif.	Penelitian ini hanya berfokus pada dua kelas, yakni sentimen positif dan negatif.	Penelitian ini menggunakan metode <i>lexicon</i> untuk pelabelan sentiment apakah termasuk	Penelitian ini menunjukkan SVM layak diimplementasikan karena akurasi yang lebih baik



No.	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
	<i>Support Vector Machine</i>	Serta, menggunakan Google Review sebagai sumber <i>dataset</i> .			positif dan negatif. Serta, metode SVM untuk proses klasifikasinya.	dari Naïve Bayes, yakni sebesar 100%.
11.	Penerapan <i>Word N-Gram</i> untuk <i>Sentiment Analysis Review</i> menggunakan Metode <i>Support Vector Machine</i> (Studi Kasus: Aplikasi Sambara)	Menggunakan metode SVM untuk proses klasifikasi ulasan.	Data ulasan yang digunakan penelitian ini bersumber dari Google Play terkait aplikasi Sambara.	Penelitian ini tidak dituliskan mengenai pelabelan <i>dataset</i> yang dilakukan. Serta, hanya berfokus pada dua kelas, yakni sentiment positif dan negatif.	Penelitian ini menggunakan metode SVM untuk proses klasifikasinya. Serta, berfokus pada penerapan <i>word n-gram</i> yang memiliki akurasi tertinggi.	Penelitian ini menunjukkan bahwa penerapan <i>word N-gram</i> dapat meningkatkan akurasi nilai akurasi jika data berjumlah banyak. Nilai akurasi tertinggi mencapai 92% dengan nilai AUC 0.954 pada penerapan <i>tri-gram</i> .
12.	Analisis Sentimen Pelanggan Toko Online JD.ID menggunakan Metode <i>Naïve Bayes Classifier</i> Berbasis Konversi Ikon Emosi	Mengekstrak informasi dari <i>text</i> dan emotikon.	Menggunakan metode <i>Naïve Bayes Classifier</i> disertai pembobotan TF-IDF. <i>Dataset</i> yang digunakan bersumber dari Twitter.	Penelitian ini berfokus terhadap opini atau penilaian pengguna di Twitter. Jumlah <i>emoticon</i> yang di <i>convert</i> sebanyak 15 emoji yang mana hanya merepresentasikan empat emosi, yakni senang,	Penelitian ini menggunakan pembobotan TF-IDF bertujuan untuk ekstraksi emosi dari <i>text</i> dan <i>convert emoticon</i> . Serta, menggunakan <i>Naïve Bayes Classifier</i> (NBC) untuk klasifikasi	Penelitian ini menunjukan akurasi model dengan <i>convert emoticon</i> memiliki akurasi yang lebih tinggi dibandingkan tanpa <i>convert emoticon</i> . Serta, sentiment masyarakat terhadap JD.ID

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
				sedih, kesal, dan bahagia.	sentiment tersebut.	bersifat positif, netral, dan negatif.

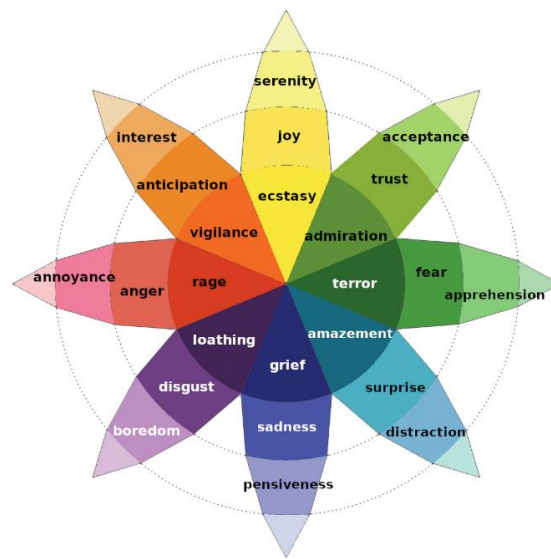
## 2.2 Landasan Teori

### 2.2.1 Analisis Emosi

Emosi adalah keadaan kompleks yang terdiri atas perasaan, pikiran, perilaku, dan reaksi fisik seseorang terhadap rangsangan internal atau eksternal. Emosi mempengaruhi kehidupan manusia sehari-hari [29]. Emosi berperan penting dalam bagaimana seseorang mengambil keputusan, memahami hal-hal di sekitarnya, dan berkomunikasi dengan orang lain [33]. Salah satu bentuk komunikasi dalam hubungan interpersonal yang penting adalah ekspresi emosi. Emosi dasar manusia terbagi menjadi dua, yaitu emosi positif dan negatif [34].

Terdapat enam emosi dasar menurut Paul Ekman, yakni marah (*anger*), gembira (*joy*), jijik (*disgust*), sedih (*sad*), terkejut (*surprise*), dan takut (*fear*) [35][36][37]. *Anger* adalah ekspresi dari rasa jengkel dan geram. *Joy* dapat didefinisikan sebagai bahagia, candu, dan ketenangan. *Disgust* adalah bentuk ekspresi dari ketidaksukaan dan benci. *Sad* adalah bentuk ekspresi kesuraman atau duka. *Surprise* dapat didefinisikan sebagai bentuk ekspresi takjub serta ketidakpastian. Serta, *fear* adalah bentuk rasa takut dan teror.

Dari enam emosi dasar tersebut menurunkan ekspresi emosi lainnya, seperti *anticipation* dan *trust* menurut Robert Plutchik [38][39]. *Anticipation* didefinisikan sebagai rasa waspada atau minat. *Trust* adalah penerimaan atau kekaguman. Gambar 2.1 menunjukkan *wheel of emotion* dari Robert Plutchik. Emosi-emosi tersebut dapat ditunjukkan baik secara langsung maupun dituangkan dalam bentuk kata serta emotikon dalam berbagai platform media sosial [40]. Kemudian, bentuk emosi tersebut dapat menjadi sumber informasi yang dapat berguna secara efektif dan efisien untuk berbagai macam kepentingan.



Gambar 2. 1. *Plutchik's wheel of emotion* [41]

Analisis emosi [29] adalah kegiatan yang mengidentifikasi dan interpretasi perasaan dan persepsi dari suatu satuan yang berwujud, seperti teks, video, atau gambar. Analisis emosi dapat dilakukan menggunakan pendekatan *lexicon*, *machine learning*, atau metode lainnya yang sesuai sebagai alat bantu [17]. Analisis emosi dengan pendekatan *lexicon* menggunakan pencarian berbasis kata kunci mencari kata kunci emosi yang diberikan pada beberapa keadaan psikologis. Dimana memiliki referensi berupa kamus kata atau *corpus* data yang mengandung label emosi untuk mengidentifikasi emosi dari objek yang dianalisis. Contoh *lexicon* adalah *EmoSenticNet* dan *NRC Emolex*, dsb. Analisis emosi dengan pendekatan *machine learning* menggunakan algoritma *machine learning*, seperti Naïve Bayes, SVM, *Decision Tree*, dsb.

### 2.2.2 *NRC Emolex*

Pendekatan berbasis *lexicon* adalah metode menggunakan kamus atau *corpus* kata [42]. Kamus tersebut dapat berisi kata-kata yang mengandung label positif dan negatif untuk sentimen ataupun label emosi yang terkait. Pendekatan *lexicon* dapat digunakan untuk menentukan apakah ulasan pengguna mengandung label tersebut. Salah satu metode *lexicon* adalah *NRC Emolex*.

*NRC Emolex* adalah kamus kata yang mencakup sekitar 14.000 istilah bahasa Inggris yang sudah menyediakan versi *lexicon* dalam lebih dari 100 bahasa, salah satunya bahasa Indonesia [18][43]. *NRC Emotion Lexicon* atau *NRC Emolex* menggunakan anotasi biner, yaitu 0 dan 1 [18]. *NRC Emolex* dapat menunjukkan ‘ya’ atau ‘tidak’ untuk kelas sentimen (positif dan negatif) dan kelas emosi. Kelas emosi pada *NRC Emolex* sesuai dengan teori emosi yang dikemukakan oleh Robert Plutchik, yaitu *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, dan *trust*. Emosi-emosi tersebut dapat dideteksi secara otomatis ketika menganalisis teks [42]. Rincian ringkasan terkait *NRC Emolex* ditunjukkan pada Tabel 2.2.

Tabel 2. 2. Rincian ringkasan tentang *NRC Emolex*

<i>Lexicon</i> asosiasi	Versi	Jumlah istilah	Kategori	Skor Asosiasi	Metode Pembuatan
<i>Lexicon</i> Asosiasi Kata-Emosi dan Kata-Sentimen					
<i>NRC Word-Emotion Association Lexicon (Emolex)</i>	0.92	14,182 unigrams (kata)	<b>sentimen:</b> positif, negatif <b>emotions:</b> marah, antisipatif, muak, takut, bahagia, sedih, terkejut, dan yakin	0 atau 1	<b>Manual:</b> Melalui <i>crowdsourcing on Mechanical Turk</i> .  <b>Domain:</b> Umum

### 2.2.3 Support Vector Machine

*Support Vector Machine* (SVM) merupakan salah satu metode *supervised learning* yang dikembangkan oleh Boser, Guyon, dan Vapnik pada tahun 1992 [44]. Metode ini sering digunakan untuk kasus klasifikasi dengan menggunakan *hyperplane* atau garis pemisah. Dalam pemilihan *hyperplane* terbaik, dilakukan dengan memaksimalkan margin. Margin merupakan jarak antara *support vector* dengan *hyperplane*, dimana *support vector* merupakan objek terluar dari setiap kelas [45]. Penelitian [21] SVM *hyperplane* diterapkan dengan rumusan (2.1).

$$w_0 + w_1x_1 + w_2x_2 = 0$$

(2. 1)

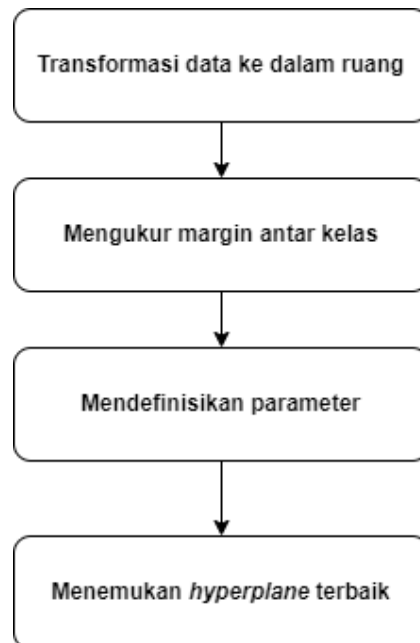
$w_0$  merupakan konstanta bias jarak vertikal antar *hyperplane* dan asal koordinat.  $w_1$  dan  $w_2$  adalah fitur  $x_1$  dan  $x_2$  dalam ruang fitur. Dimana  $x_1$  dan  $x_2$  adalah fitur dari sampel data yang ingin diprediksi. Setiap titik yang berada di atas *hyperplane* memenuhi rumusan yang ditentukan (2.2).

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (2.2)$$

Setiap titik yang berada di bawah *hyperplane* juga memenuhi rumusan yang ditentukan (2.3).

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (2.3)$$

Kelebihan dari metode SVM adalah dapat melakukan klasifikasi dengan jumlah data yang sedikit [45]. Selain itu, metode ini dapat digunakan dengan mudah pada *dataset* karena Python memiliki *library scikit learn*. Metode SVM juga bisa menerapkan pemisahan linear pada input data *non-linear* dengan menggunakan kernel. Gambar 2.2 menunjukkan alur dari SVM.



Gambar 2. 2. Tahapan SVM [45]

#### 2.2.4 Analisis Persepsi

Persepsi merupakan ekspresi pengalaman seseorang mengenai objek, peristiwa, atau hubungan yang diperoleh melalui kesimpulan informasi dan interpretasi pesan. Persepsi bisa didefinisikan sebagai tindakan atau pandangan seseorang tentang suatu peristiwa atau kejadian yang terjadi [46]. Analisis persepsi adalah mengetahui bagaimana seorang individu memandang atau menilai suatu objek. Analisis persepsi ini berguna untuk memahami bagaimana individu membentuk pandangan mereka terhadap suatu hal.

Analisis persepsi dari ulasan pengguna dapat menggunakan teknik *word cloud*. Teknik ini digunakan sebagai representasi grafis dengan menghitung *term frequency* jumlah kata yang muncul pada setiap kelompok kata [25]. *Word cloud* menyoroti kata-kata yang paling sering ada di dalam ulasan dan membantu untuk mengidentifikasi kata-kata kunci yang dominan. Ukuran kata menunjukkan seberapa pentingnya kata tersebut.

#### 2.2.5 Visualisasi Word Cloud

Teknik *word cloud* digunakan untuk representasi grafis dari teks ulasan. *Word cloud* memfokuskan kata-kata yang paling sering digunakan dalam ulasan dan membantu untuk mengidentifikasi kata-kata kunci yang dominan secara mendalam [47]. Metode ini memvisualisasikan data teks dimana kata-kata yang paling sering digunakan ditampilkan dengan ukuran yang lebih besar dan lebih menonjol, sedangkan kata-kata yang muncul lebih jarang ditampilkan dengan ukuran yang lebih kecil [25]. Ulasan masing-masing dari setiap objek museum akan disajikan dalam bentuk *word cloud*. Hal tersebut dapat memberikan gambaran yang menunjukkan persepsi pengunjung terhadap objek-objek museum.

Berikut adalah langkah-langkah umum untuk membuat visualisasi *word cloud*.

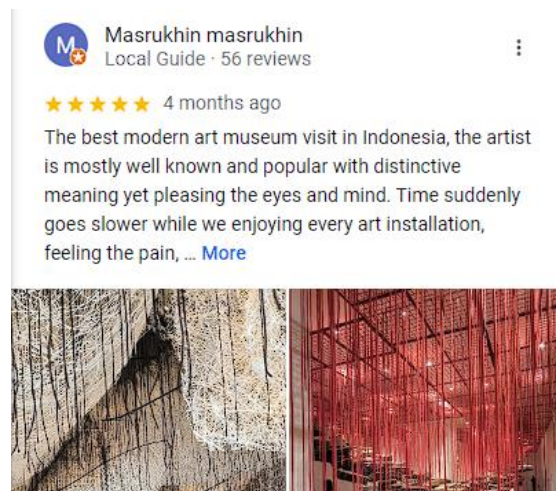
1. Menghitung frekuensi kata: Hitung frekuensi kata dalam teks. Tahap ini dihitung berapa kali setiap kata muncul dalam teks.
2. Mengatur ukuran kata: Atur ukuran kata berdasarkan frekuensi relatifnya.

Kata-kata yang paling sering digunakan akan memiliki ukuran yang lebih besar dan lebih menonjol.

3. Visualisasi *word cloud*: Pada tahapan ini dilakukan pembuatan visualisasi *word cloud* menggunakan *library* atau alat visualisasi data yang tersedia, seperti *matplotlib* dan *seaborn*. *Library* tersebut menyediakan fungsi untuk membuat visualisasi *word cloud*.
4. Penyesuaian dan interpretasi: Tahap terakhir adalah melakukan penyesuaian tampilan *word cloud*, seperti warna, font, atau penyesuaian lainnya. Setelah itu, dilakukan interpretasi *word cloud* dengan melihat kata-kata yang berukuran lebih besar untuk mengetahui kata-kata yang paling sering digunakan dalam ulasan.

#### **2.2.6 Ulasan Google Maps**

Pada Google Maps terdapat beberapa fitur salah satunya adalah Google Maps Reviews. Google Maps adalah sebuah platform yang disediakan oleh Google terkait informasi suatu tempat, menghitung rute antar lokasi, menawarkan peta dan foto, serta ulasan dari suatu tempat. Pengguna dapat memberikan ulasan dari tempat yang telah dikunjungi pada Google Maps Reviews [48]. Oleh karena itu, informasi yang disediakan oleh Google Maps menarik minat para peneliti walaupun Google Maps tidak dirancang dan dikembangkan dari sudut penelitian [49]. Gambar 2.3 menunjukkan ulasan pada Google Maps.



Gambar 2. 3. Ulasan Google Maps

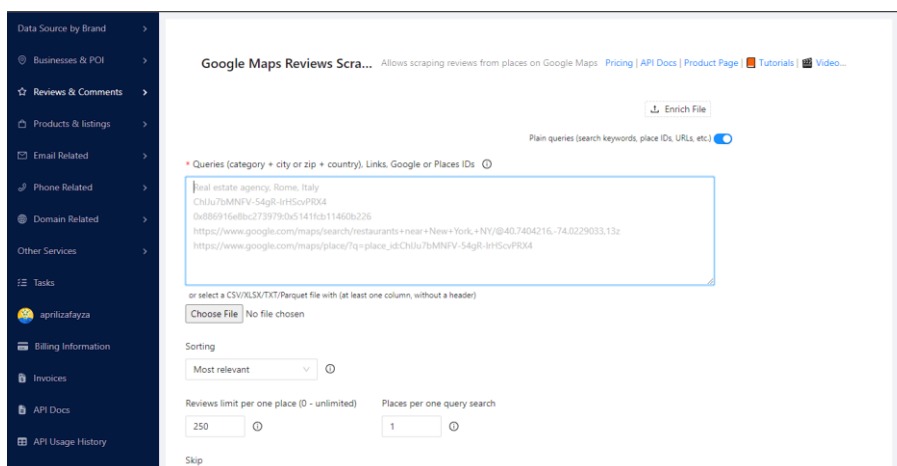
### 2.2.7 *Web Scraping*

*Web scraping* adalah salah satu teknik mengumpulkan data dari halaman *web* di internet dengan menggunakan API (*Application Programming Interface*). *Web Scraping* digunakan untuk mendapatkan data atau informasi dari sebuah situs *web* yang berupa teks, tautan, video, audio, atau dokumen secara otomatis [50]. *Scraping* bisa dilakukan dengan metode manual ataupun secara otomatis menggunakan *tools*. Salah dua *scraping tool* adalah Outscrapper dan Instant Data Scraper.

Outscrapper merupakan *scrape tool* yang menyediakan data yang cepat dan andal. Outscrapper memiliki antarmuka yang sederhana dan intuitif [51]. Platform ini dapat *scraping* data dari Google Maps, ulasan Google Maps, penelusuran Google, ulasan Google Play, komentar YouTube, dan sebagainya. Kelebihan Outscrapper dalam melakukan *scraping* data adalah keandalan, menggunakan teknologi mutakhir, pembaruan reguler, kecepatan, *cloud scraping*, dan tanpa batas. Selain itu, data yang telah di-*scrape* dapat disimpan ke dalam berbagai format, seperti XML dan CSV.

Salah satu layanan Outscrapper adalah Google Maps Reviews Scraper. Layanan ini dapat digunakan secara gratis untuk 500 ulasan pertama. Tampilan halaman layanan Google Maps Reviews Scraper ditunjukkan pada Gambar 2.4.





Gambar 2. 4. Tampilan Google Maps Reviews Scraper

Tabel 2.3. menjelaskan *data dictionary*, yaitu nama kolom dan deskripsi yang dapat diperoleh saat melakukan *scraping* menggunakan layanan ini.

Tabel 2. 3. *Review data dictionary*

<i>Column Name</i>	<i>Description</i>
<i>name</i>	Nama tempat di Google Maps.
<i>google_id</i>	<i>Unique identifier</i> dari tempat
<i>location_link</i>	Tautan tempat
<i>reviews_link</i>	Tautan ke ulasan tempat tersebut.
<i>reviews_per_score</i>	Objek JSON dengan skor dan jumlah ulasan untuk setiap skor.
<i>rating</i>	<i>Rating</i> dari ulasan.
<i>review_id</i>	<i>Unique identifier</i> dari ulasan.
<i>author_link</i>	Tautan ke profil <i>author</i> di Google Maps.
<i>author_title</i>	Nama penulis.
<i>author_id</i>	<i>Unique identifier</i> dari penulis.
<i>author_image</i>	Tautan gambar dari halaman profil penulis.
<i>review_text</i>	Teks ulasan.
<i>review_img_url</i>	Tautan gambar dari ulasan.
<i>owner_answer</i>	Teks balasan dari pemilik.
<i>owner_answer_timestamp</i>	<i>Timestamp</i> ketika pemilik memberikan balasan.
<i>owner_answer_timestamp_datetime_utc</i>	<i>Datetime</i> ketika pemilik memberikan balasan.
<i>review_link</i>	Tautan dari ulasan
<i>review_rating</i>	<i>Rating</i> dari ulasan.

<i>Column Name</i>	<i>Description</i>
<i>review_timestamp</i>	<i>Timestamp</i> ketika ulasan dibuat.
<i>review_datetime_utc</i>	<i>Datetime</i> ketika ulasan dibuat.
<i>review_likes</i>	Jumlah suka dari ulasan
<i>reviews_id</i>	<i>Unique identifier</i> dari ulasan

Instant Data Scraper merupakan *tools* ekstraksi data secara otomatis dari situs halaman *web* [52]. Data yang berhasil dikumpulkan dapat diekspor sebagai file excel atau csv. Instant Data Scraper menggunakan kecerdasan buatan *heuristic* dari struktur HTML untuk memprediksi data mana yang paling relevan pada halaman HTML. Jika prediksinya tidak memuaskan maka pengguna dapat menyesuaikan pilihan. Teknologi *scraping* ini lebih nyaman karena tidak memerlukan *library* skrip *scraping* yang dibuat secara manual oleh pengguna.

Instant Data Scraper dapat digunakan secara gratis. *Tools* ini dapat digunakan bersama alat SEO, sistem perekrut CRM, alat manajemen *lead* penjualan, atau kampanye pemasaran email. *Web scraping* dan mengunduh data hasil *scraping* menjadi mudah serta *tools* ini memiliki keamanan dan privasi data karena data yang diambil tidak meninggalkan jejak *browser* pengguna. Instant Data Scraper memiliki beberapa fitur sebagai berikut.

1. Mendeteksi data untuk ekstraksi dengan kecerdasan buatan.
2. Mendeteksi kapan data dinamis telah dimuat.
3. Penyesuaian waktu tunda dan waktu maksimal menunggu untuk kecepatan *scraping* yang diinginkan.
4. Adanya dukungan untuk paginasi halaman di situs *web*.
5. Navigasi otomatis ke halaman berikutnya.
6. Adanya dukungan untuk gulir tak terbatas.
7. Pratinjau data yang berhasil di-*scraping* dengan dukungan *copy* dan *paste*.
8. Ekspor data sebagai file excel atau csv.
9. Pembaruan nama kolom dan penyaringan data yang diekstrak.

### 2.2.8 Data Preprocessing

*Preprocessing* adalah proses transformasi *raw text* menjadi masukan dan melakukan beberapa tahapan dasar untuk mengubah atau menghapus data *noise* yang tidak digunakan dalam pengolahan data selanjutnya [24][53]. Berikut tahapan *data preprocessing* yang biasa dilakukan sebagai berikut.

1. *Case folding* merupakan proses konversi seluruh huruf besar menjadi *lowercase* atau huruf kecil. Contohnya “AkU PErGi” menjadi “aku pergi”.
2. *Convert emoticon* adalah tahap menerjemahkan arti emosi yang terkandung dari emotikon.
3. *Remove punctuation and number* adalah proses menghapus tanda baca, angka, dan simbol yang tidak diperlukan.
4. *Convert slangword* adalah langkah mengubah kata yang tidak baku atau tidak resmi menjadi kata baku dalam Bahasa Indonesia yang merujuk pada Kamus Baku Bahasa Indonesia (KBBI).
5. *Stopword removal* adalah proses menghapus kata-kata yang tidak berhubungan dalam menentukan inti suatu dokumen dan paling sering muncul dalam dokumen, seperti “serta”, “sedang”, “merupakan”, “ialah”, dalam dokumen berbahasa Indonesia.
6. *Stemming* adalah tahap perubahan kata menjadi kata dasar dengan menghapus sufiks atau prefiks pada kata. Contoh kata “melukis” menjadi “lukis” dan “menyapu” menjadi “sapu”. Pada penelitian ini tahapan *stemming* tidak dilakukan karena kata baik dengan sufiks ataupun prefiks memiliki makna emosi yang berbeda.

### 2.2.9 TF-IDF

*Term Frequency – Invers Document Frequency* atau TF-IDF digunakan untuk pembobotan kata dimana mengubah setiap kata menjadi bentuk numerik. Skor TF-IDF menunjukkan seberapa penting atau relevan suatu *term* atau kata dalam dokumen tertentu. *Term Frequency* (TF) artinya jika kata tertentu sering muncul di dalam semua dokumen, maka kata tersebut memiliki tingkat

kepentingan yang tinggi dibandingkan kata-kata lain yang muncul lebih sedikit. Nilai TF diperoleh menggunakan rumus (2.4).

$$TF_{t,d} = \frac{f_{t,d}}{\sum t,d} \quad (2.4)$$

$f_{t,d}$  adalah frekuensi setiap kata (t) yang muncul di dalam kalimat. Nilai tersebut dibagi dengan total keseluruhan kata (t) yang terdapat di dalam kalimat (d) yang ditunjukkan pada notasi  $\sum t,d$ . *Invers Document Frequency* (IDF) artinya jika kata tertentu sering muncul di dalam sebuah dokumen tetapi juga sering muncul di beberapa dokumen lain, maka kata tersebut dapat dianggap tidak penting. Nilai IDF diperoleh menggunakan rumus (2.5).

$$IDF_{t,d} = \log \frac{|D|}{df(t) + 1} \quad (2.5)$$

$|D|$  merupakan jumlah dokumen yang ada di dalam koleksi dibagi dengan jumlah dokumen di mana muncul kata (t) yang dinotasikan  $df(t)$ . Penambahan angka 1 bertujuan untuk menghindari angka 0 jika  $df(t)$  tidak ditemukan pada *corpus*. Rumus (2.6) digunakan untuk menghitung nilai TF-IDF menggunakan hasil perhitungan sebelumnya.

$$W_{t,d} = TF_{t,d} \times IDF_t \quad (2.6)$$

### 2.2.10 Seleksi Fitur dengan *Chi-Squared*

Untuk meningkatkan kinerja model, mengurangi *overfitting*, dan mempercepat proses pelatihan, sangat penting untuk memilih *subset* fitur yang paling informatif dan relevan ketika mengembangkan model prediktif. *Dataset* sering mengandung banyak fitur yang mungkin tidak memberikan informasi yang berharga atau bahkan mengandung *noise* sehingga tahapan seleksi fitur menjadi sangat penting untuk dilakukan. Salah satu metode seleksi fitur yang digunakan adalah *SelectKBest*, yang ada di *library scikit-learn*. Metode ini akan memilih sejumlah  $k$  fitur terbaik berdasarkan skor atau nilai tertentu, termasuk statistik *chi-squared* [54]. Dalam penelitian ini, statistik *chi-squared* digunakan untuk mengukur signifikansi hubungan antara data fitur dan data target.

Kemunculan fitur dan kategori adalah dua peristiwa dalam seleksi fitur *chi-squared* yang didasarkan pada teori statistika [55]. Kemudian, berdasarkan perhitungan pada persamaan 2.7, setiap nilai *term* diurutkan dari yang tertinggi.

$$X^2 = \sum_{i=1}^m \frac{(O_i - E_i)^2}{E_i} \quad (2.7)$$

Di mana  $O_i$  adalah frekuensi observasi untuk fitur  $i$  pada kelas target tertentu,  $E_i$  adalah frekuensi harapan untuk fitur  $i$  pada kelas target tertentu dengan  $m$  adalah jumlah kelas target.

### 2.2.11 *Principal Component Analysis (PCA)*

*Principal Component Analysis*, juga dikenal sebagai PCA, merupakan penggunaan sesedikit mungkin komponen untuk mengurangi dimensi data yang digunakan. Karl Pearson pertama kali menggunakan teknik ini pada tahun 1901. Mengurangi dimensi data tetapi tetap mempertahankan variansi tertinggi adalah konsep utama PCA. Tujuan PCA untuk menemukan variabel baru yang signifikansi setelah mengurangi ukuran *dataset* terlebih dahulu [56]. Seringkali, PCA melakukan pemangkasan dengan menyimpan hanya komponen utama dan mengabaikan komponen lainnya. Dua alasan utama untuk pemangkasan di antaranya: pertama, pemangkasan dapat membantu mengurangi kelebihan data. Jika dua komponen utama dapat menangkap sebagian besar perbedaan, *dataset* dapat dianggap kurang lebih dua dimensi. Kedua, pemangkasan membantu menunjukkan tingkat kompleksitas data. Untuk menghilangkan pengaruh dimensi dan urutan besaran, matriks data asli matriks dinormalisasi melalui metode *Z-Score* [57]. Normalisasi ini dinyatakan pada persamaan 2.8.

$$Z = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (2.8)$$

di mana  $\bar{x}_j$  adalah jumlah rata-rata sampel ke- $j$ , dan  $S_j$  adalah standar deviasi dari sampel ke- $j$ .

### 2.2.12 *Resampling*

Data yang tidak seimbang merupakan masalah di mana jumlah kelas dari satu *dataset* berbeda secara signifikan [58]. Dalam hal ini, kelas minoritas terdiri atas kelas yang memiliki jumlah sampel sedikit dan kelas mayoritas terdiri atas kelas yang memiliki jumlah sampel yang besar. *Dataset* yang tidak seimbang ini dapat mengakibatkan performa model cenderung condong ke kelas mayoritas. Contohnya, pada kelas mayoritas model memiliki akurasi yang baik, tetapi memiliki akurasi yang rendah pada kelas minoritas. Penanganan terhadap data yang tidak seimbang dalam klasifikasi data terdapat tiga pendekatan umum, yaitu *pre-processing methods*, pendekatan berbasis algoritma, dan pendekatan *hybrid*.

Pada pendekatan *pre-processing method* terdapat beberapa metode, seperti *over-sampling*, *under-sampling*, atau menggabungkan kedua metode tersebut untuk mendapatkan jumlah sample yang kurang lebih sama di setiap kelas. *Over-sampling* bertujuan untuk meningkatkan ukuran kelas minoritas agar mendapatkan kelas yang seimbang. *Under-sampling* bertujuan untuk menggunakan set acak dari sampel kelas mayoritas untuk menyeimbangkan kelas, sementara sisa sampelnya diabaikan. Metode *hybrid* adalah metode yang menggunakan kedua teknik *resampling* tersebut, di mana untuk mendapatkan data pelatihan yang seimbang dilakukan *under-sampling* untuk menghapus instans yang tidak memiliki informasi yang berguna, kemudian *over-sampling* dilakukan untuk menduplikasi instans yang sudah ada.

SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan salah satu contoh metode *hybrid* [59]. SMOTE menghasilkan contoh sintentis dalam kelas minoritas dan merupakan alternatif untuk duplikasi sampel kelas minoritas. SMOTE termasuk ke dalam kategori *hybrid* karena menggunakan kombinasi *under-sampling* dan *over-sampling* untuk mengatasi masalah ketidakseimbangan distribusi data dan tidak hanya mengandalkan *under-sampling* saja. Tahapan *resampling* dengan SMOTE [60] diidentifikasi sebagai berikut.

1. Mengambil fitur a yang sedang diperiksa dari kelas minoritas dan menemukan k fitur terdekat dari fitur b dalam kelas minoritas yang sama
2. Untuk setiap fitur b yang dipilih, hitung selisih antara fitur a dan b, kemudian kalikan selisihnya dengan angka acak antara 0 dan 1, serta tambahkan ke fitur b yang sedang diperiksa.

### 2.2.13 Evaluasi Model

*Confusion matrix* merupakan matriks  $n \times n$  yang digunakan untuk mengevaluasi kinerja model klasifikasi, dimana  $n$  adalah jumlah kelas [33][45]. Tabel 2.4 menunjukkan contoh *confusion matrix*  $2 \times 2$  yang terdiri atas empat metrik utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Tabel 2. 4. *Confusion matrix*

Kelas Prediksi	Kelas Aktual	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Positive</i>
Negatif	<i>False Negative</i>	<i>True Negative</i>

Penjelasan metrik utama berdasarkan Tabel 2.4 sebagai berikut

1. *True Positive (TP)*: Total data positif yang diprediksi dengan benar sebagai positif.
2. *True Negative (TN)*: Total data negatif yang diprediksi dengan benar sebagai negatif.
3. *False Positive (FP)*: Total data negatif yang salah diprediksi sebagai positif.
4. *False Negative (FN)*: Total data positif yang salah diprediksi sebagai negatif.

Berdasarkan penjelasan metrik utama, nilai TP dan TN adalah ketika data diprediksi dengan benar, sedangkan FP dan FN adalah ketika data salah diprediksi. Dari metrik evaluasi tersebut dapat dihitung akurasi, presisi, *recall*,

dan *F1-score*. Akurasi menunjukkan seberapa akurat model klasifikasi yang dibangun dalam memprediksi keseluruhan data. Nilai akurasi didapatkan dari rumus (2.9).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9)$$

Presisi menunjukkan kemampuan model klasifikasi yang dibangun dalam mengidentifikasi data positif dengan benar. Nilai presisi didapatkan dari rumus (2.10).

$$Precision = \frac{TP}{TP + FP} \quad (2.10)$$

*Recall* menunjukkan seberapa baik model klasifikasi yang dibangun dalam mengidentifikasi semua data positif sebenarnya. Nilai *recall* didapatkan dari rumus (2.11).

$$Recall = \frac{TP}{TP + FN} \quad (2.11)$$

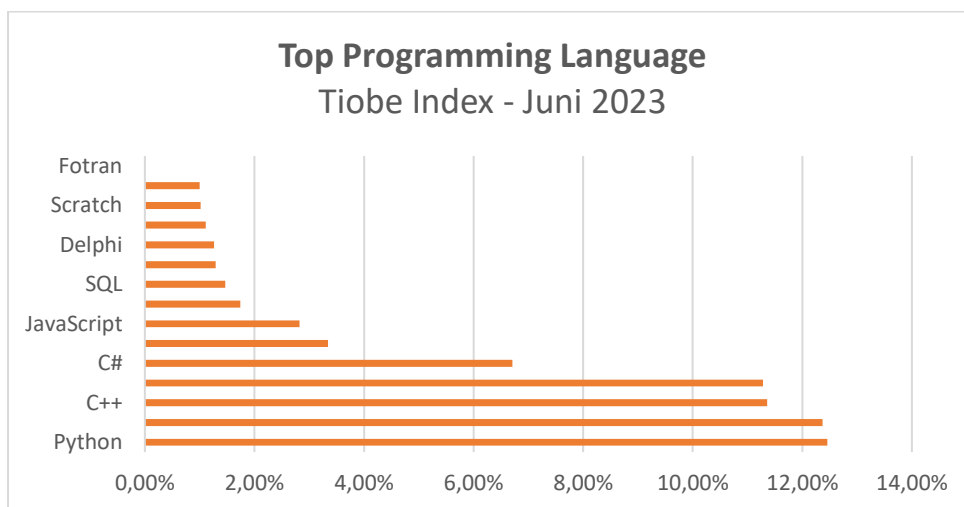
*F1-score* merupakan nilai rata-rata harmonik antara presisi dan *recall*. *F1-score* didapatkan dari rumus (2.12).

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (2.12)$$

#### 2.2.14 Python

Python adalah salah satu bahasa pemrograman tingkat tinggi atau *high level language* [61]. Menurut Tiobe, Python merupakan bahasa pemrograman paling populer pada Juni 2023 yang ditunjukkan pada Gambar 2.5. Python bersifat *multi-purpose* karena dapat digunakan untuk *web development*, *data analyst*, *data visualization*, *task automation*, dan *machine learning*. Program dengan Python dijalankan melalui interpreter, artinya tidak memerlukan komputasi terlebih dahulu ketika menjalankannya. Hal ini membuat Python sangat mudah dipelajari.





Gambar 2. 5. Peringkat kepopuleran bahasa pemrograman menurut Tiobe Juni 2023 [62]

Pada tahun 1985 Guido van Rossum mengembangkan Python untuk pertama kalinya. Python masih terus dikembangkan hingga saat ini [61][63]. Saat ini terdapat dua jenis versi Python, yakni versi 2.x dan 3.x. Python memiliki perbaikan yang paling drastis pada versi 3.x, yaitu adanya dukungan *unicode* yang lebih baik dan aspek *command* utama telah disesuaikan sehingga memberikan kemudahan bagi pemula. Serta, lebih konsisten. Python versi 3.x memiliki performa yang lebih baik dibandingkan 2.x.

Secara umum Python memiliki kelebihan dibandingkan bahasa pemrograman lainnya sebagai berikut.

- a. Python mendukung Pemrograman Berbasis Objek (OOP).
- b. Python dapat diinstal di beberapa *operation system*, yaitu Windows, Mac OSX Unix/Linux, iOS, iOS, dan solaris (*Multi platform*).
- c. Bersifat *open source*.
- d. Kemudahan memahami struktur dan perintah Python.
- e. Tersedia banyak modul atau *library* yang dapat digunakan untuk menyelesaikan pekerjaan tertentu.
- f. Dapat diintegrasikan dengan beberapa DBMS yang populer, seperti MySQL, Oracle, dll.
- g. *Script* Python dapat digunakan secara *portable*.

- h. Python mendukung pengembangan pemrograman berbasis GUI (*Graphical User Interface*).
- i. Python dapat diintegrasikan dengan program yang dibuat dengan bahasa pemrograman lain.
- j. Mendukung pemrograman *socket*.
- k. Kode program Python dapat dikompilasi menjadi *byte-code* untuk aplikasi skala besar.

### 2.2.15 Jupyter Notebook

Jupyter Notebook adalah salah satu aplikasi *web open-source* gratis yang memungkinkan pengguna untuk menulis dokumen yang berisi teks penjelasan, persamaan, visualisasi, serta kode yang terbuka dan hasil eksekusi mereka. Jupyter Notebook saat ini menjadi sangat populer [64]. Jupyter Notebook digunakan sebagai aplikasi untuk membuat notebook berbasis Python dan *libraries* lainnya, khususnya kepentingan *data science* [61]. Notebook tersebut dapat dipergunakan untuk media presentasi, laporan penelitian, ataupun mengajar. Jupyter Notebook ini tersedia dalam Anaconda dan Win Python. Gambar 2.6 menunjukkan tampilan awal Jupyter Notebook.



Gambar 2. 6. Tampilan *document root* Jupyter Notebook