

BAB 2

DASAR TEORI

2.1 KAJIAN PUSTAKA

Penelitian Miranti Alysha Zulia Larasati, Nurul Anisa Sri Winarsih, Muhammad Syaifur Rohman dan Galuh Wilujeng Saraswati pada tahun 2022 dengan judul “Penerapan Metode *K-Means Clustering* dalam Menganalisis Sentimen Masyarakat terhadap K-Poppers pada *Twitter*”. Penelitian ini bermaksud untuk menganalisis sentimen dari masyarakat Indonesia atas keberadaan *K-Pop* pada media sosial *twitter*. Dataset yang diambil sebanyak 1000. Penelitian ini memanfaatkan algoritma *k-means*. Hasil pengujian data memperlihatkan metode *k-means clustering* baik dalam menyelesaikan klasifikasi sentimen pada media sosial *twitter* dengan hasil nilai *silhouette index* sebesar 0,687974. Hasil dapat dipengaruhi oleh data uji dan data latih yang memiliki kualitas baik. Sehingga, ketika dilakukan percobaan terhadap data yang lain dapat mempengaruhi hasil klasifikasi [11].

Penelitian tentang “Analisis Sentimen *Twitter* terhadap Penggunaan Artis Korea Selatan sebagai *Brand Ambassador* Produk Kecantikan Lokal” oleh Ristyani Slamet, Windu Gata, Annisa Novtariany, Khairunisa Hilyati dan Febri Ainun Jariyah pada tahun 2022. Tujuan dilakukannya penelitian ini adalah untuk menghasilkan model algoritma klasifikasi terbaik. Selain itu, penelitian ini bisa dimanfaatkan untuk pengambilan keputusan *brand skincare* lokal yang memilih artis korea sebagai *brand ambassador*. Dataset diambil dari sumber data *twitter* dengan mencari kata kunci *skincare* dan BA pada bulan Maret 2022 hingga April 2022. Algoritma yang dimanfaatkan pada penelitian ini adalah SVM (*Support Vector Machine*) dan *Naïve Bayes*. Output yang dihasilkan adalah perbandingan performansi dari kedua algoritma yang dipakai. Proses pelabelan kata pada penelitian ini dilakukan secara manual yang mengacu pada *dictionary* berkonotasi negatif dan positif. Hasil pengujian menunjukkan bahwa model terbaik untuk analisis sentimen menggunakan algoritma SVM dengan nilai akurasi sebesar 83.60%, precision sebesar 83.86% dan recall sebesar 99.62%. Hasil analisis

klasifikasi sentimen produk kecantikan local dengan penggunaan artis korea selatan sebagai *brand ambassador* menunjukkan kecenderungan masyarakat dengan sentimen positif. [12].

Sofianita Mutalib dan Nurzeatul Hamimah Abdul Hamid pada tahun 2018 melakukan penelitian yang berjudul “*Web Mining in Classifying Youth Emotions*”. Penelitian ini bertujuan untuk mempelajari emosi anak muda di dua negara berbeda yaitu Malaysia dan Korea terpaut dengan tren KPop. Analisis tweet pada penelitian ini dilakukan dengan mengekstraksi 6.000 tweet. Klasifikasi emosi dibagi menjadi 6 kelas yaitu *joy, sadness, fear, anger, disgust and surprise* dengan menggunakan algoritma Naïves bayes. [13].

Dessy Angelina, Umi Hayati dan Gifthera Dwilestari pada tahun 2023 melakukan penelitian dengan judul “Penerapan Metode *Support Vector Machine* pada Sentimen Analisis Pengguna Twitter Terhadap Konser K-Pop”. Penelitian ini bertujuan untuk untuk mengelompokkan sentimen pengguna twitter terhadap konser kpop sesuai dengan pola permasalahan dengan memanfaatkan Algoritma Support Vector Machine (SVM). Analisis tweet pada penelitian ini dilakukan dengan mengekstraksi 841 tweet yang sudah bersih. Klasifikasi dibagi menjadi 2 kelas yaitu positif dan negatif [14]. Tabel 2.1 menunjukkan penelitian terdahulu.

Tabel 2. 1 Penelitian Terdahulu.

No	Penulis (Tahun)	Metode	Hasil	Perbedaan Penelitian
1	Miranti Alysha Zulia Larasati, Nurul Anisa Sri Winarsih, Muhammad Syaifur Rohman, Galuh Wilujeng Saraswati (2022)	<i>Metode K-Means Clustering</i>	Hasil penelitian ini menunjukkan metode <i>K-Means clustering</i> baik dalam melakukan klasifikasi sentimen masyarakat pada media sosial <i>twitter</i> . Nilai <i>silhouette index</i> yang dihasilkan sebesar 0,687974.	Menggunakan algoritma <i>Naïve Bayes</i> dan dataset diambil pada tahun 2023

No	Penulis (Tahun)	Metode	Hasil	Perbedaan Penelitian
2	Ristyani Slamet, Windu Gata, Annisa Novtariany, Khairunisa Hilyati, Febri Ainun Jariyah (2022)	Metode Penelitian Yang Digunakan Yaitu Metode KDD (<i>Knowledge Discovery In Database</i>) Menggunakan Algoritma <i>Support Vector Machine</i> (SVM) Dan <i>Naïve Bayes</i>	Hasil penelitian ini adalah klasifikasi sentimen dua kelas yaitu positif dan negatif. Model terbaik didapatkan oleh algoritma SVM dengan nilai akurasi 83,60%, precision 83,86% dan recall 99,62%. Sedangkan dengan <i>Naïve Bayes</i> , nilai Akurasi 82,65%, precision 87,81% dan recall 92,11%.	Sentimen yang digunakan dibagi menjadi 3 kelas yaitu positif, netral dan negatif dan dataset diambil pada tahun 2023.
3	Zura Izlita Razak, Shuzlina Abdul-Rahman, Sofianita Mutalib and Nurzeatul Hamimah Abdul Hamid (2018)	Metode <i>Naïve Bayes classifier</i>	Hasil dari penelitian ini memperlihatkan bahwa emosi tertinggi dari 6 kelompok orang yang melakukan <i>update</i> tentang KPop adalah ' <i>Joy</i> ' dengan jumlah sebanyak lebih dari 250 tweet. Sedangkan emosi terendah pada penelitian ini yaitu ' <i>Disgust</i> ' dengan jumlah tweet terendah di 10.	Menampilkan performasi dari algoritma <i>naïve bayes</i> dan kelas sentimen dibagi menjadi 3 kelas yaitu positif, netral, dan negatif.

No	Penulis (Tahun)	Metode	Hasil	Perbedaan Penelitian
4	Dessy Angelina, Umi Hayati dan Gifthera Dwilestari (2023)	<i>Support Vector Machine</i>	Hasil penelitian ini menghasilkan performa sistem dalam menghasilkan ketepatan analisis sentiment yang berbentuk confusion matrix . Hasil akurasi sebesar 76,64%. tingkat error pada penelitian ini sebesar 23,36% dengan micro average 23,35%.	Sentimen yang digunakan dibagi menjadi 3 kelas yaitu positif, netral dan negatif. Dan menggunakan algoritma <i>Naïve Bayes</i> .

2.2 DASAR TEORI

2.2.1 KPop

KPop merupakan singkatan dari *Korean Pop*. Kata ini mengarah pada musik populer yang berasal dari Korea Selatan dengan berbagai *genre*. KPop juga mencorakkan istilah untuk menggambarkan jenis musik, lagu, dan tarian yang dipabrikasi oleh perusahaan idol Korea Selatan. Umumnya, idol KPop tidak sekedar musisi dan penari berbakat, namun juga idol dengan karakter fisik yang menarik dan menawan. Banyak artikel yang hanya membahas sudut pandang negatif dari datangnya budaya KPop di Indonesia. Sebagaimana fanatisme, kurang cinta budaya asli dan lainnya. Namun, sungguh banyak hal positif dari datangnya KPop. Banyak generasi yang mengikuti fenomena KPop, mereka tentu mengagumi artis Korea dengan alasan yang kuat. Salah satu sebab yang paling umum adalah idola KPop mencerminkan sikap disiplin dan pekerja keras. Hal ini tampak dari fakta bahwa artis KPop harus melewati tahap karantina atau pelatihan terlebih dahulu sebelum mereka akhirnya dapat debut. Setiap idol berlatih sesuai waktu yang telah ditentukan oleh aturan perusahaan yang menaungi mereka. Oleh

karena itu, tentunya perjuangan para calon idol menjadi contoh ke penggemar untuk selalu disiplin, rajin dan semangat dalam mencapai cita-citanya [15]. Perkembangan musik KPop bermula saat Korea Selatan berada pada masa di bawah penjajahan Jepang. kemajuannya dimulai dengan gelombang budaya *Korean pop* secara meluas ke seluruh negara Asia. kendatipun, seiring berjalannya waktu, mulai menyebar ke wilayah Eropa bahkan Amerika. Beberapa orang yakin bahwa "KPop merupakan hobi yang sehat karena memberikan kesempatan untuk bertemu teman-teman dari seluruh dunia. Selain itu, memberikan penggemar kesempatan untuk mengerti *genre* musik dari perspektif yang berbeda-beda dan belajar untuk tidak menilai buruk selera orang lain". Sebagai *genre*, KPop memiliki banyak perbedaan faktor untuk dikonsumsi secara luas. Salah satu faktornya ialah penggunaan bahasa Inggris pada lirik, judul, dan nama. Sebenarnya, tidak hanya bahasa Inggris namun dicampur dengan bahasa Korea. Bahasa Inggris dipakai untuk mengekspresikan globalisasi dan modernitas. Bahasa Korea untuk menyatakan lokalitas, konservatisme atau orientalisme. Idol grup Korea, kerap kali terdiri atas beberapa anggota dari berbagai negara. Grup biasanya dapat dikompilasi dengan anggota Korea, Cina, Jepang, Amerika, Thailand, bahkan Indonesia. Sehingga, memungkinkan grup untuk mempunyai jalan masuk secara langsung untuk dapat berpromosi di negara-negara luar Korea. Unsur-unsur *global* ini tentu menyebabkan banyak yang mengkritik *genre* ini sudah kurang keaslian budayanya. Memang benar bahwa dengan sengaja mengglobal adalah sebuah ciri khas kesuksesan KPop [16].

2.2.2 Twitter

Sejalan dengan meningkatnya penggunaan media sosial saat ini, persentase pengguna *Twitter* di Indonesia juga tumbuh pesat dari tahun ke tahun. Pada tahun 2019, jumlah pengguna *Twitter* di Indonesia mencapai 6,43 juta atau sekitar 52% dari keseluruhan pengguna media sosial. *Twitter* merupakan salah satu platform media sosial yang telah dikenal digunakan oleh instansi pemerintah, sektor swasta, dan masyarakat sipil. *Twitter* juga digunakan sebagai penghubung media yang signifikan menyampaikan informasi dan situasi terkini karena bisa menjangkau dengan lebih cergas dan efektif. *Twitter* merupakan salah satu layanan jaringan sosial yang dioperasikan oleh *twitter, Inc.* *Twitter* sebagai media

dapat menggabungkan komunikasi antar hubungan dengan pengguna lain yang memungkinkan informasi dikumpulkan dari berbagai sumber [17]. *Twitter* merupakan salah satu media sosial yang memiliki fitur *retweet* yang dapat digunakan oleh setiap penggunanya untuk mengunggah kembali informasi atau *tweet* yang memungkinkan penyebaran informasi pada media sosial *twitter* menjadi lebih cepat. *Twitter* juga merupakan media sosial yang dapat digunakan untuk analisis sentimen dengan menggunakan data *tweet* yang diperoleh dengan melakukan *crawling* data [18]. *Twitter* adalah sumber yang kaya data karena orang-orang melakukan *tweet* setiap saat. Analisis sentimen *Twitter* adalah salah satu teknik untuk menentukan agregat perasaan orang dari data *microblogging* opini. Ini memainkan peran penting dalam mengidentifikasi sentimen atau pendapat individu dan dampaknya terhadap masyarakat [19].

2.2.3 Analisis Sentimen

Sikap, perasaan, pandangan, serta pendapat merupakan bagian yang penting dalam menganalisis perilaku seseorang yang dapat dirujuk sebagai sentimen. Analisis sentimen juga diketahui sebagai *opinion mining* atau pengkajian opini, yang berkaitan dengan pemeriksaan opini yang diarahkan ke entitas manapun [20]. Analisis sentimen adalah salah satu kunci pengembangan *Artificial Intelligence* (AI). Analisis sentimen dapat dianggap sebagai masalah klasifikasi teks karena prosesnya mencakup beberapa operasi yang berakhir dengan mengklasifikasikan apakah teks tertentu mengekspresikan sentimen positif, negatif dan lainnya. Saat ini, analisis sentimen sangat dikenal, tidak hanya di kalangan peneliti, tetapi juga perusahaan, pemerintah, dan organisasi. Meningkatnya penggunaan internet telah membuat web menjadi sumber informasi universal dan terpenting. Jutaan orang mengungkapkan pendapat dan sentimen mereka di forum, blog, wiki, jejaring sosial, dan sumber daya web lainnya. Pendapat dan sentimen tersebut sangat relevan dengan kehidupan kita sehari-hari, sehingga analisis data digunakan untuk memantau opini publik dan membantu pengambilan keputusan. Misalnya, postingan *Twitter* telah digunakan untuk memprediksi sesuatu. Untuk alasan ini, bidang analisis sentimen mendapatkan lebih banyak minat dalam satu setengah dekade terakhir di kalangan komunitas

riset. Sejak tahun 2004, analisis sentimen telah menjadi bidang penelitian yang paling cepat berkembang dan paling aktif, karena telah terjadi peningkatan besar-besaran [21].

2.2.4 Data Mining

Data mining dikenal juga dengan nama KDD (*Knowledge Discovery in Database*). Tetapi, pada tahun 1995, diselenggarakan *International KDD Conference* di Montreal. Hal itu berhasil memutuskan bahwa KDD ialah suatu proses mengenali informasi berguna serta mengenali pola yang dapat dimengerti dari data. Tujuan utama proses KDD ialah memprediksi nilai-nilai penting dari variabel-variabel yang ada atau dari kelompok data yang dapat diinterpretasikan oleh manusia. Sesuai dengan tujuan tersebut, maka proses dalam mengenali informasi baru dan penemuan pola tersebut perlu diaplikasikan dengan *data mining*. *Data mining* ialah salah satu aspek yang banyak didukung oleh cabang ilmu lain di dalam teknologi informasi. Seperti statistik, *machine learning*, *basis data*, sistem pakar, algoritma, visualisasi data, dan lainnya [22]. *Data mining* merupakan salah satu kegiatan pada di bagian *software* yang bisa menciptakan ROI (*Return of Investment*) yang cukup besar. [23].

2.2.5 Pre-Processing

Pre-processing data merupakan proses pembersihan teks untuk klasifikasi. Teks *online* biasanya mengandung cukup banyak *noise* serta bagian yang tidak informatif seperti link, simbol dan iklan. Sebelum diproses data perlu dikurangi *noise* sehingga dapat membantu kinerja *classifier* dan mempercepat dilakukannya proses klasifikasi dalam analisis sentimen. [24].

2.2.5.1 Cleaning

Cleaning ialah tahap pembersihan data dari karakter-karakter serta kata yang tidak diperlukan. Tujuan *cleaning* adalah untuk mengurangi *noise* yang terkandung dalam teks sehingga perhitungan dalam prosesklasifikasi tidak optimal. Pada tahap ini dilakukan pembersihan simbol dan karakter yang tidak diperlukan. Selain itu dilakukan pula perubahan simbol tertentu yang berhubungan dengan kata kunci pencarian yaitu format *space*, yang dalam hal ini

space akan berubah menjadi simbol + (*plus*), sehingga dilakukan pengubahan kembali menjadi *space*.

2.2.5.2 Case Folding

Case Folding ialah tahap mengubah data menjadi format huruf yang sama. Tujuannya untuk mengurangi redundansi data sehingga proses perhitungan klasifikasi menjadi optimal. Contohnya adalah mengganti format teks menjadi *lowercase* atau *uppercase* sesuai kebutuhan dalam proses klasifikasi.

2.2.5.3 Tokenizing

Tokenizing ialah proses memisahkan data menjadi kata-perkata berdasarkan delimiter yang digunakan. Pada tahap ini, dilakukannya pemecahan suatu gabungan kata atau disebut juga kalimat menjadi persatu kata untuk mempermudah pada proses selanjutnya [25].

2.2.6 Pelabelan Data

Pada analisis sentimen, pelabelan merupakan tahap yang dilakukan setelah proses *preprocessing*. Fungsi dari pelabelan data adalah mengklasifikasikan kelas pada data [26]. Data yang dihasilkan pada proses pelabelan akan digunakan untuk melatih sistem untuk mengenali model atau sebagai data *training* [27].

2.2.7 Term Frequency - Inverse Document Frequency (TF-IDF)

Term weighting atau pembobotan kata dimaksudkan untuk memberi nilai bobot pada setiap kata. Untuk menghitung bobot ini, diperlukan dua hal yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* adalah jumlah kata atau *term* tertentu yang ada pada suatu dokumen. *Inverse Document Frequency* ialah frekuensi kemunculan kata atau *term* di semua dokumen. Nilai IDF akan berbanding terbalik dengan total dokumen yang mengandung *term* tertentu. Di semua dokumen, *Term* yang jarang muncul memiliki nilai IDF yang lebih tinggi daripada nilai IDF *term* yang sering muncul. Jika pada setiap dokumen berisi *term* tertentu, maka nilai IDF *term* adalah 0. Rumus TF-IDF adalah sebagai berikut [28] :

$$W_{dt} = tf_{dt} \times \log \left(\frac{N}{df_t} \right) \quad (2.5)$$

Dimana :

W_{dt} = bobot *term* ke-t terhadap dokumen d

tf_d = jumlah kemunculan term t dalam dokumen - d

N = jumlah dokumen secara keseluruhan

df_t = jumlah dokumen yang mengandung *term* t

2.2.8 Klasifikasi

Klasifikasi merupakan tahapan untuk membuat model yang menjelaskan konsep atau kelas data. Tujuannya untuk memprediksi kelas dari suatu objek yang labelnya tidak diketahui. Model bisa berupa aturan “jika maka”, *decision tree* atau formula matematis. Beberapa macam Metode *classification* adalah *Bayesian*, *Neural Network*, *Genetic Algorithm*, *Fuzzy*, *Case-based Reasoning* dan *K-Nearest Neighbor* [29]. Klasifikasi ialah cara mengelompokkan benda berdasar karakter yang ada pada objek klasifikasi. Dalam prosesnya, klasifikasi bisa dilakukan menggunakan banyak cara. Baik secara manual atau dengan bantuan teknologi. Klasifikasi yang dilakukan secara manual ialah klasifikasi yang dilakukan manusia tanpa bantuan algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan memanfaatkan teknologi mempunyai beberapa algoritma, seperti *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, *Fuzzy* serta Jaringan Saraf Tiruan. Klasifikasi adalah salah satu hal penting dalam data mining. klasifikasi dibuat dari kumpulan data latih dengan kelas yang telah ditentukan sebelumnya. Proses akan mengelompokkan fitur ke dalam kelas yang sesuai. Vektor fitur dari data latih telah diketahui kelasnya terlebih dahulu. Kemudian, vektor fitur data latih tersebut digunakan untuk merancang suatu ketetapan model. pola ini disebut terbimbing *supervised*. Definisi lain dari klasifikasi teks ialah diberikannya sebuah masukan berupa dokumen (d) dan terdapat sebuah kelas yang sudah tetap contohnya $C = \{c_1, c_2, \dots, c_3\}$. Nantinya, output yang dihasilkan adalah sebuah model prediksi kelas. Beberapa contoh dari klasifikasi teks adalah deteksi spam, analisis sentiment, identifikasi penulis, identifikasi usia atau gender, identifikasi bahasa, dan lainnya [30].

2.2.9 Naïve Bayes

Naïve Bayes merupakan algoritma kuat yang digunakan untuk mengklasifikasikan data berdasarkan probabilitas. Algoritma ini hanya bekerja

pada teorema *Bayes* dan menggunakan berbagai probabilitas untuk mengklasifikasikan data. Di kelas *Naïve Bayes* probabilitas maksimum dianggap sebagai kelas yang diprediksi. *Naïve Bayes* juga dikenal sebagai *Maximum a Posterior Naïve Bayes* memiliki berbagai kelebihan dan kekurangan domain yang berbeda. Algoritma ini, merupakan algoritma yang cepat dan sangat skalabel. Algoritma *Naïve Bayes* juga digunakan pada *Multiclass* dan *Binary Classification*. Selain itu, algoritma ini juga dapat digunakan pada dataset kecil dan dengan demikian juga memberikan hasil yang baik [31]. Pada Teorema *Bayes*, jika terdapat dua kejadian terpisah (A dan B), maka Teorema *Bayes* dirumuskan [32]:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A) \quad (2.1)$$

Di mana:

A = Data dengan kelas yang belum diketahui

B = Hipotesa data B merupakan suatu kelas spesifik

P(A|B) = Probabilitas hipotesis A berdasarkan kondisi B

P(A) = Probabilitas hipotesis A

Teorema bayes sesekali dioptimalkan dengan hukum probabilitas total, menjadi seperti berikut [32]:

$$P(A|B) = \frac{P(A)P(B|A)}{\sum_{i=1}^n P(A|B)} \quad (2.2)$$

Untuk menerangkan teorema *Naïve Bayes*, perlu diperhatikan bahwa proses klasifikasi memerlukan beberapa syarat untuk menentukan kelas mana yang sesuai untuk sampel yang dianalisis. Oleh karena itu, teorema *Bayes* disesuaikan sebagai berikut [32]:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2.3)$$

Variabel C mewakili kelas, sedangkan variabel F1...Fn mewakili karakteristik petunjuk yang diperlukan untuk menyelesaikan klasifikasi. Maka dari itu, rumus tersebut kemudian menjelaskan bahwa peluang masuknya suatu sampel dengan karakteristik tertentu ke dalam kelas C (*posterior*) sama dengan probabilitas munculnya kelas C (sebelum sampel masuk, seringkali disebut *prior*), dikalikan dengan probabilitas kemunculan karakteristik sampel di kelas C (disebut juga *likelihood*), dibagi dengan probabilitas karakteristik sampel yang muncul

secara global (disebut juga *evidence*). Oleh karena itu, rumus (2.3) juga dapat ditulis secara sederhana sebagai berikut [32]:

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (2.4)$$

Nilai *evidence* selalu sama untuk setiap kelas dalam satu sampel. Nilai dari Posterior sebelumnya kemudian dipadankan dengan nilai berikutnya dari Posterior kelas lain untuk menentukan di kelas mana suatu sampel akan ditempatkan [30].

2.2.10 Evaluasi

Pada proses Evaluasi, akan diterapkan *confusion matrix* untuk menghitung *precision*, *recall*, dan *accuracy* yang bertujuan untuk mengetahui keakuratan dari pemodelan yang telah diterapkan pada data latih [33]. Contoh *confusion matrix* klasifikasi data dalam 3 kelas dalam bentuk Tabel 2.2 [34]:

Tabel 2. 2 Multiclass Confusion Matrix.

		Diklasifikasikan Sebagai		
		Positif	Netral	Negatif
Aktual	Positif	<i>True Positif</i> (A)	<i>False Netral</i> (B)	<i>False Negatif</i> (C)
	Netral	<i>False Positif</i> (D)	<i>True Netral</i> (E)	<i>False Negatif</i> (F)
	Negatif	<i>False Positif</i> (G)	<i>False Netral</i> (H)	<i>True Negatif</i> (I)

Untuk perhitungan menggunakan persamaan sebagai berikut:

2.2.10.1 Accuracy

Accuracy adalah rasio dari total prediksi klasifikasi yang teridentifikasi dengan benar. *Accuracy* dapat dihitung dengan cara membagi total data prediksi benar dengan jumlah keseluruhan data uji. Seperti ditunjukkan pada persamaan berikut ini [34] :

$$Accuracy = \frac{True\ Positif\ (A) + True\ Negatif\ (I) + True\ Netral\ (E)}{jumlah\ data\ tes} \times 100\% \quad (2.6)$$

2.2.10.2 Precision

Precision adalah ketepatan jumlah yang diinginkan dengan hasil prediksi yang diberikan. *Precision* positif dapat dihitung dengan cara membagi total data

prediksi benar bernilai positif dibagi total data prediksi benar bernilai positif yang ditambah dengan total data prediksi yang salah bernilai positif. Seperti ditunjukkan pada persamaan berikut ini:

$$Precision\ Positif = \frac{True\ Positif\ (A)}{True\ Positif\ (A)+False\ Positif\ (D)+False\ Positif\ (G)} \quad (2.7)$$

$$Precision\ Negatif = \frac{True\ Negatif\ (I)}{True\ Negatif\ (I)+False\ Negatif\ (C)+False\ Negatif\ (F)} \quad (2.8)$$

$$Precision\ Netral = \frac{True\ Netral\ (E)}{True\ Netral\ (E)+False\ Netral\ (B)+False\ Netral\ (H)} \quad (2.9)$$

$$Precision = \frac{Precision(Positif)+Precision(Netral)+Precision(Negatif)}{Total} \times 100\% \quad (2.10)$$

2.2.10.3 Recall

Recall menggambarkan suksesnya model klasifikasi dapat menemukan informasi yang tepat. Penghitungan *recall positif* ialah rasio data prediksi benar bernilai positif dibagi dengan data prediksi benar bernilai positif ditambah data prediksi salah bernilai negatif dan netral. Seperti ditunjukkan pada persamaan berikut ini:

$$Recall\ Positif = \frac{True\ Positif\ (A)}{True\ Positif\ (A)+False\ Netral\ (B)+False\ Negatif\ (C)} \quad (2.11)$$

$$Recall\ Negatif = \frac{True\ Negatif\ (I)}{True\ Negatif\ (I)+False\ Netral\ (H)+False\ Positif\ (G)} \quad (2.12)$$

$$Recall\ Netral = \frac{True\ Netral\ (E)}{True\ Netral\ (E)+False\ Negatif\ (D)+False\ Positif\ (F)} \quad (2.13)$$

$$Recall = \frac{Recall(Positif)+Recall(Netral)+Recall(Negatif)}{Total} \times 100\% \quad (2.14)$$

2.2.10.4 F-1 Score

F-1 *Score* adalah parameter tunggal yang mengukur seberapa sukses *recall* dan *precision* ketika digabungkan. Hasil dari F1-*Score* merupakan perkalian antara *precision* dan *recall* dibagi *precision* ditambah *recall* didapat dari perhitungan hasil *precision* dan *recall* dibagi dengan hasil penjumlahan *precision* dan *recall* kemudian dikalikan dua dan perhitungannya menggunakan persamaan [35] :

$$f - Score\ Positif = 2 \frac{Precision\ Positif \times Recall\ Positif}{Precision\ Positif + Recall\ Positif} \quad (2.15)$$

$$f - Score\ Negatif = 2 \frac{Precision\ Negatif \times Recall\ Negatif}{Precision\ Negatif + Recall\ Negatif} \quad (2.15)$$

$$f - Score\ Netral = 2 \frac{Precision\ Netral \times Recall\ Netral}{Precision\ Netral + Recall\ Netral} \quad (2.15)$$

$$f - Score = \frac{f-Score(Positif)+f-Score(Netral)+f-Score(Negatif)}{Total} \times 100\% \quad (2.15)$$