

Analisis Sentimen Masyarakat Indonesia Terhadap Saham GOTO Group Pada Media Sosial Twitter Menggunakan Metode *Naive Bayes Classifier* dan *Support Vector Machine*

Ichwanul Kahfi Prasetya^[1] dan Adatul Mukarromah^[2]

Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember

Jalan Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: ichwanul.18062@mhs.its.ac.id^[1] dan adatul@statistika.its.ac.id^[2]

Abstrak—Pasar Modal memiliki peran penting dalam perekonomian suatu negara. Instrumen pasar modal yang paling populer adalah saham yang banyak dipilih para investor karena kemampuannya dalam memberikan tingkat keuntungan yang menarik. *Initial Public Offering* (IPO) dapat didefinisikan sebagai proses transformasi suatu perusahaan menjadi perusahaan publik dengan menjual sebagian atau seluruh perusahaan. GOTO Group adalah ekosistem digital terbesar di Indonesia yang terdiri dari berbagai layanan melalui platform Gojek, Tokopedia, dan GOTO Financial. GOTO Group telah mengumumkan rencana melaksanakan IPO di Bursa Efek Indonesia dengan nilai mencapai Rp 18 triliun. Dengan latar belakang perusahaan ini yang berbasis pada teknologi membuat informasi IPO ini menyebar luas secara cepat dengan adanya teknologi informasi dan internet. Twitter merupakan media sosial yang menjadi wadah berdiskusi dan bertukar pendapat atau opini melalui *tweet* antar sesama penggunanya. Opini yang disampaikan bisa berupa opini positif maupun negatif yang kemudian dapat dilakukan analisis sentimen. Analisis sentimen merupakan salah satu teknik untuk mengekstrak sebuah informasi berupa sikap seseorang terhadap suatu isu atau kejadian dengan mengelompokkan polaritas dari sebuah teks. Pada penelitian ini metode yang digunakan adalah *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM). Perbandingan metode tersebut menggunakan *10-fold cross validation* dengan kriteria kebaikan klasifikasi *accuracy* menunjukkan bahwa hasil performa metode SVM Kernel RBF lebih baik daripada metode NBC dan SVM Kernel Linier dalam mengklasifikasikan data teks sentimen masyarakat terhadap Saham GOTO Group.

Kata Kunci—GOTO, *Naive Bayes Classifier*, Sentimen, *Support Vector Machine*, Twitter

I. PENDAHULUAN

Pasar Modal memiliki peran penting dalam perekonomian suatu negara. Instrumen pasar modal yang paling populer adalah saham (*stocks*) yang banyak dipilih para investor karena kemampuannya dalam memberikan tingkat keuntungan yang menarik. Saham dapat didefinisikan sebagai tanda kepemilikan atau penyertaan modal seseorang atau pihak dalam suatu perusahaan. Pada sisi lain, menerbitkan saham merupakan salah satu pilihan perusahaan ketika memutuskan untuk pendanaan perusahaan [1]. Proses awal penerbitan saham disebut dengan *Initial Public Offering* (IPO). IPO dapat didefinisikan sebagai proses transformasi suatu perusahaan menjadi perusahaan publik dengan menjual sebagian atau seluruh perusahaan melalui penawaran pasar perdana [2].

GOTO Group adalah ekosistem digital terbesar di Indonesia yang terdiri dari layanan transportasi *on-demand*, *e-commerce*, pengiriman makanan dan bahan makanan, logistik dan pemenuhan, serta layanan keuangan dan pembayaran melalui platform Gojek, Tokopedia, dan GOTO Financial [3]. GOTO Group telah mengumumkan rencana melaksanakan IPO di Bursa Efek Indonesia (BEI).

Teknologi informasi memegang peranan kehidupan manusia karena dapat menghubungkan dan menyajikan berbagai informasi melalui internet. Pada internet terdapat berbagai macam blog pribadi, situs jejaring sosial, dan *microblog* lainnya yang memuat fakta maupun opini [4]. Twitter sebagai media sosial sering menjadi wadah berdiskusi dan bertukar opini melalui *tweet* antar sesama penggunanya [5]. Opini yang disampaikan bisa berupa opini positif maupun negatif yang kemudian diubah menjadi data untuk dilakukan analisis sentimen.

Analisis sentimen merupakan salah satu teknik untuk mengekstrak sebuah informasi berupa sikap seseorang terhadap suatu isu atau kejadian dengan mengelompokkan polaritas dari sebuah teks [6]. Metode klasifikasi yang terdapat pada ranah ilmu statistika bermacam-macam. Pada penelitian ini metode yang digunakan adalah *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).

Permasalahan pada penelitian ini adalah mengetahui opini masyarakat Indonesia terhadap saham GOTO Group (\$GOTO) berupa analisis sentimen pada media sosial Twitter menggunakan metode NBC dan SVM. Melalui penelitian ini diharapkan dapat membantu membantu pihak-pihak yang ingin mengetahui opini masyarakat serta kata kunci yang identik terhadap saham emiten GOTO Group dengan analisis sentimen melalui *tweet* pengguna Twitter.

II. TINJAUAN PUSTAKA

A. Sentiment Analysis

Sentiment analysis mengacu pada bidang yang luas dari *text mining* yang bertujuan menganalisis pendapat, sikap, emosi, dan penilaian seseorang terhadap suatu topik, produk, organisasi, individu, ataupun kegiatan tertentu lainnya [6]. Tugas dasar analisis sentimen adalah mengklasifikasikan teks yang ada dalam sebuah dokumen dan kemudian menentukan pendapat yang diungkapkan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral [7].

B. Text Mining

Text mining merupakan salah satu cabang ilmu *data mining* yang dapat digunakan untuk menganalisis suatu data dalam bentuk dokumen teks. *Text mining* adalah proses melakukan analisis pada suatu data teks dengan mencari pola-pola informasi dari teks yang tidak terstruktur untuk mendapatkan informasi berkualitas [8].

C. Confix-Stripping Stemmer

Stemmer merupakan aplikasi untuk memotong imbuhan pada suatu kata yang dijalankan menggunakan algoritme tertentu. Algoritma *stemming* dengan Bahasa Indonesia telah dikembangkan oleh Nazief dan Adriani yang kemudian dikenal dengan *Nazief and Adriani's Stemmer*. Algoritma *Confix-Stripping Stemmer* merupakan pengembangan dari algoritma sebelumnya yaitu *Nazief and Adriani's Stemmer*. Algoritma ini dikembangkan dengan perbaikan algoritma menyesuaikan kaidah Bahasa Indonesia dengan tujuan untuk meningkatkan hasil *stemming* yang diperoleh [9].

D. Term Frequency Inverse Document Frequency

Term Frequency Inverse Document Frequency (TF-IDF) merupakan sebuah metode pembobotan dalam prosesi ekstraksi data teks. Tujuan dari TF-IDF adalah untuk menemukan jumlah kata yang diketahui (*TF*) setelah dikalikan dengan beberapa banyak *tweet* dimana suatu kata tersebut muncul (*IDF*). Rumus untuk menemukan pembobot dengan TF-IDF ditunjukkan pada persamaan (1).

$$t_{nm} = tf_{nm} \times idf \quad (1)$$

$$idf = \log \left(\frac{M}{df_{nm}} \right)$$

Keterangan:

t_{nm} = bobot TF-IDF pada kata kunci ke- n dan *tweet* ke- m
 TF_{nm} = jumlah kemunculan kata kunci ke- n pada *tweet* ke- m
 DF_{nm} = jumlah *tweet* ke- m yang mengandung kata kunci n
 M = jumlah seluruh *tweet*

E. K-Fold Cross Validation

K-fold cross validation merupakan salah satu metode yang digunakan untuk membagi data menjadi dua bagian yaitu data *training* dan data *testing*. *K-fold cross validation* secara iteratif membagi data menjadi data *training* dan *testing* (Gokgoz & Subasi, 2015). Metode ini diawali dengan membagi data menjadi K subset atau *folds* yang saling bebas, dengan masing-masing *fold* S_1, S_2, \dots, S_k berukuran sama. Data *training* dan *testing* dilakukan dengan iterasi sebanyak K kali. Pada iterasi ke- k , *fold* S_k dengan $1 \leq k \leq K$ digunakan sebagai *testing set*, dan partisi yang tersisa digunakan kolektif untuk *training* model, dilakukan hingga setiap *fold* telah digunakan sebagai data *testing* dan *training* [10].

F. Naive Bayes Classifier

Teorema *Bayes* merupakan teorema yang mengacu pada konsep probabilitas bersyarat dua kejadian *random* [11]. Secara umum teorema *Bayes* dapat ditunjukkan pada persamaan (2).

$$P(A | B) = \frac{P(A)P(B | A)}{P(B)} \quad (2)$$

Naive Bayes Classifier (NBC) merupakan salah satu teknik *data mining* yang sering digunakan untuk melakukan klasifikasi data dalam jumlah yang besar dan dapat untuk memprediksi probabilitas keanggotaan suatu kelas [12]. Kelebihan NBC adalah algoritmanya sederhana tetapi memiliki akurasi yang tinggi. Klasifikasi *Naive Bayes* yang merujuk pada teorema *bayes* mempunyai persamaan (3).

$$P(y_k | t_1, t_2, \dots, t_M) = \frac{P(t_1, t_2, \dots, t_M | y_k)P(y_k)}{P(x_1, x_2, \dots, x_M)} \quad (3)$$

Keterangan:

y_k = sentimen *tweet*; k = positif, negatif
 t_m = bobot TF-IDF pada kata kunci ke- m ; $m = 1, 2, \dots, M$
 $P(y_k)$ = probabilitas kategori *tweet*
 $P(t_m)$ = probabilitas kondisi bobot TF-IDF pada kata kunci
 $P(y_k | t_m)$ = probabilitas kategori *tweet* berdasarkan kondisi bobot TF-IDF
 $P(t_m | y_k)$ = probabilitas kondisi bobot TF-IDF berdasarkan kondisi kategori *tweet*

Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi yang kemudian dilakukan klasifikasi data pada kategori yang paling tepat [13]. Berikut merupakan tahapan algoritma dari metode pengklasifikasian *naive bayes*.

1. Variabel t merupakan sekumpulan dokumen bobot kata kunci yang direpresentasikan dengan atribut t_m untuk $m=1, 2, \dots, M$ dimana M adalah jumlah seluruh kata kunci. sedangkan y_k adalah himpunan kategori *tweet*.
2. Menghitung nilai $P(y_k)$ pada data *training* menggunakan persamaan (4) berikut.

$$P(y_k) = \frac{|train_k|}{|train|} \quad (4)$$

Keterangan:

$P(y_k)$ = probabilitas *tweet* berkategori k
 $train_k$ = jumlah *tweet* data *training* berkategori k
 $train$ = jumlah *tweet* data *training*

Setiap probabilitas kata setiap kategori dihitung saat *training* pada persamaan (5).

$$P(t_m | y_k) = \frac{t_{mk} + 1}{\sum_{m=1}^M t_{mk} + \sum_{m=1}^M \sum_{k=-1,1} t_{mk}} \quad (5)$$

Keterangan:

t_{mk} = bobot TF-IDF pada kata kunci ke- m berkategori k ;
 $m = 1, 2, \dots, M$; $k = -1, 1$
 $\sum t_{mk}$ = jumlah bobot TF-IDF dengan kategori k
 $\sum \sum t_{mk}$ = jumlah seluruh bobot TF-IDF

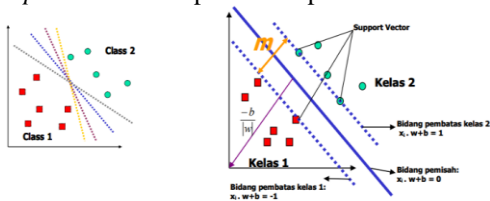
3. Melakukan klasifikasi ke dalam kelompok kategori dengan menghitung probabilitas tertinggi kata t_m pada kategori y_k menggunakan persamaan (6).

$$Y_{MAP} = \arg \max_{y_k = Y} P(y_k) \prod_{m=1}^M P(t_m | y_k) \quad (6)$$

G. Support Vector Machine

Support Vector Machine (SVM) merupakan metode statistika yang mempelajari area yang memisahkan antar kategori menggunakan suatu bidang [14]. Tujuan dari metode ini adalah membangun pemisah optimum sehingga dapat digunakan untuk proses klasifikasi. SVM dikembangkan dengan prinsip *linier classifier*. Namun dalam kasus nyata sering dijumpai data yang tidak linier sehingga dikembangkan SVM untuk kasus *nonlinier* dengan konsep kernel.

SVM pada *linearly separable* data adalah penerapan metode SVM pada data yang dapat dipisahkan secara linier. Fungsi dari *hyperplane* itu digunakan sebagai pemisah dua buah kelas pada input space yang sering disimbolkan dengan -1 dan +1. Misalkan x_n adalah dataset *tweet* untuk $n=1,2,...,N$ dan $y_k \in \{+1, -1\}$ adalah label kelas dari data x_n . Penggambaran *linearly separable data* dapat dilihat pada Gambar 1.



Gambar 1 lustrasi *Hyperplane* Alternatif (Kir) dan Optimum (Kanan)

Pada Gambar 1, kedua kelas data dapat dipisahkan oleh sepasang bidang pembatas yang sejajar. Data yang berada pada bidang pembatas disebut dengan *support vector*. $-b/w$ merupakan jarak bidang pemisah yang tegak lurus dari titik pusat koordinat dan $|w|$ adalah jarak *Euclidean* dari w . Persamaan *hyperplane* dapat ditulis pada persamaan (7).

$$\begin{aligned} \vec{x}_n \cdot \vec{w} + b &\geq 0 ; \text{ untuk } y_n = +1 \\ \vec{x}_n \cdot \vec{w} + b &\leq 0 ; \text{ untuk } y_n = -1 \end{aligned} \quad (7)$$

Keterangan:

\vec{w} = normal bidang

b = posisi bidang relatif terhadap pusat koordinat (bias)

\vec{x}_n = Nilai x (*tweet*) pada baris ke- n ; $n = 1, 2, \dots, N$

y_n = Kategori y pada baris ke- n

Kemudian mencari nilai w dan b dengan persamaan (8), sehingga kelas dari suatu data *testing* (x_{new}) dapat ditemukan menggunakan persamaan (9).

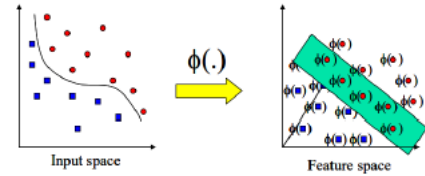
$$\begin{aligned} w &= \sum_{m=1}^M \alpha_m y_m t_m \\ \hat{b} &= 1 - w^T x \end{aligned} \quad (8)$$

$$f(t_{new}) = \text{sign}(\vec{t}_{new} + \hat{b}) \quad (9)$$

Pada kasus nyata, sangat jarang dijumpai masalah yang bersifat *linier separable*. Sebagai metode yang dikembangkan untuk kasus linier, SVM membutuhkan fungsi yang mampu membuat pemisah yang tidak linier. Fungsi yang sering digunakan untuk mengatasi hal tersebut adalah fungsi kernel. Klasifikasi data yang tidak dapat dipisahkan secara linier memerlukan modifikasi pada formula SVM agar dapat menemukan solusinya. Oleh karena itu, kedua bidang pembatas pada persamaan (7) harus diubah sehingga lebih fleksibel untuk kondisi tertentu dengan penambahan variabel *soft margin hyperlane* (ξ), sehingga menjadi persamaan (10).

$$\begin{aligned} x_n \cdot w + b &\geq 1 - \xi ; \text{ untuk kelas 1} \\ x_n \cdot w + b &\leq -1 - \xi ; \text{ untuk kelas 2} \end{aligned} \quad (10)$$

Penyelesaian data jenis ini dilakukan dengan menerapkan metode kernel. Metode kernel bekerja dengan mentransformasi data ke dalam dimensi ruang fitur sehingga data dapat dipisahkan secara linier pada *feature space*.



Gambar 2 Transformasi Vector Input ke *Feature Space*

Tahap penyelesaian data dipetakan dengan menggunakan fungsi pemetaan (transformasi) $x_n \rightarrow \phi(x_n)$ ke dalam *feature space* sehingga terdapat bidang pemisah yang dapat memisahkan data sesuai kelasnya seperti pada Gambar 2. Kemudian kelas dari suatu data *testing* (x_{new}) dapat ditemukan menggunakan persamaan (11).

$$\begin{aligned} f(x_{new}) &= \sum_{n=1}^N \alpha_n \phi(x_{n_1}) \phi(x_{n_2}) + b \\ &= \sum_{n=1}^N \alpha_n \exp(-\gamma \|x_{n_1} - x_{n_2}\|^2) + b \end{aligned} \quad (11)$$

Syarat sebuah fungsi untuk menjadi fungsi kernel adalah memenuhi teorema *Mercer* yang menyatakan bahwa matriks kernel yang dihasilkan harus bersifat *positive semi-definite*. Fungsi kernel yang umum digunakan dapat dilihat pada Tabel 1 sebagai berikut.

Tabel 1 Fungsi Kernel Pada SVM

Fungsi Kernel	Rumus $K(x_{m1}, x_{m2})$	Parameter
Linier	$x_{n_1} x_{n_2} + C$	C
RBF	$\exp(-\gamma \ x_{n_1} - x_{n_2}\ ^2)$	γ dan C

H. Evaluasi Ketepatan Klasifikasi

Pengukuran ketepatan klasifikasi dilakukan untuk melihat performa klasifikasi yang telah dilakukan. Dalam mengukur ketepatan klasifikasi perlu diketahui jumlah pada setiap kelas prediksi dan kelas aktual yang terdiri dari TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*). Keempat nilai tersebut dapat dimuat melalui *confusion matrix* pada Tabel 2 sebagai berikut.

Tabel 2 *Confusion Matrix*

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

Pengukuran yang sering digunakan untuk menghitung ketepatan klasifikasi adalah akurasi, *specificity*, dan *sensitivity* dengan data yang *balanced* pada setiap kategorinya [12]. Akurasi merupakan persentase dari total dokumen yang teridentifikasi secara tepat dalam proses klasifikasi. Berikut

merupakan rumus dalam menghitung akurasi yang ditunjukkan pada Persamaan (12).

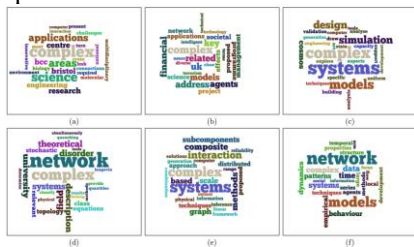
$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Sedangkan pada data yang *imbalanced*, pengukuran ketepatan klasifikasi yang digunakan adalah *Area Under Curve (AUC)*. *AUC* merupakan indikator performa kurva ROC (*Receiver Operating Characteristic*) yang dapat diringkas kinerja menjadi *classifier* menjadi satu nilai [15]. Berikut adalah rumus untuk menghitung *AUC* yang ditunjukkan pada Persamaan (13).

$$AUC = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (13)$$

I. Word Cloud

Word cloud merupakan suatu metode visualisasi dokumen teks dengan *plotting* kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi. Frekuensi dari kata yang sering muncul ditunjukkan melalui ukuran huruf kata tersebut. Contoh dari visualisasi dokumen teks dengan *word cloud* dapat dilihat pada Gambar 3.



Gambar 3 Visualisasi Data dengan *Word Cloud*

J. Twitter

Twitter merupakan layanan jejaring sosial dan *microblog* daring yang memungkinkan penggunaanya untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan kicauan (*tweet*). Kelebihan Twitter dibandingkan dengan media social lainnya yaitu mempunyai fitur *trending topic*, yaitu tema bahasan yang sedang menjadi tren di Twitter [16]. Selain memiliki berbagai fitur untuk penggunaanya, Twitter juga menyediakan fitur yang membantu pihak lain yaitu yang dapat digunakan untuk berbagai macam kepentingan yaitu API (*Application Programming Interface*). Twitter API dapat digunakan oleh para *developer* untuk mengambil data dari twitter dan mengolahnya. Twitter API terus berkembang dengan menambahkan tingkat akses tambahan bagi para peneliti dan *developer* untuk meningkatkan kemampuan dalam mengambil dan mengolah data pada Twitter [16].

K. GOTO Group

GOTO Group adalah ekosistem digital terbesar di Indonesia. Ekosistem GOTO terdiri dari makanan dan bahan makanan, logistik dan pemenuhan, serta layanan keuangan dan pembayaran melalui platform Gojek, Tokopedia, dan GOTO Financial [3]. GOTO Group memiliki 100 juta pengguna aktif bulanan, total USD 22 miliar *Gross Transaction Value (GTV)*, dan lebih dari 1,8 miliar transaksi per tahun 2020 [17]. GOTO

Group telah mengumumkan rencana melaksanakan *Initial Public Offering* (IPO) di Bursa Efek Indonesia dengan nilai valuasi mencapai Rp 18 triliun.

III. METODOLOGI

Sumber data yang digunakan pada penelitian adalah kumpulan *tweet* sentimen mengenai pendapat masyarakat terhadap saham emiten GOTO Group. Data yang digunakan berupa *tweet* dengan kata kunci ‘GOTO’ menggunakan metode *web crawling* dengan bantuan Twitter API. Data *tweets* diambil pada bulan Maret sampai Mei 2022 dengan jumlah sekitar 4950 *tweets*. Data yang digunakan terdiri dari *tweets* dengan kata kunci ‘GOTO’, klasifikasi sentimen, serta kata kunci yang terbentuk dari tahap *preprocessing*. Struktur data pada penelitian ini dapat dilihat pada Tabel 3 sebagai berikut.

Tabel 3 Struktur Data Penelitian

<i>Tweet</i> (<i>x</i>)	Klasifikasi Sentimen (<i>y</i>)	Kata Kunci (<i>t</i> ₁)	Kata Kunci (<i>t</i> ₂)	...	Kata Kunci (<i>t</i> _{<i>M</i>})
<i>x</i> ₁	<i>y</i> ₁	<i>t</i> _{1,1}	<i>t</i> _{2,1}	...	<i>t</i> _{<i>M</i>,1}
<i>x</i> ₂	<i>y</i> ₂	<i>t</i> _{1,2}	<i>t</i> _{2,2}	...	<i>t</i> _{<i>M</i>,2}
<i>x</i> ₃	<i>y</i> ₃	<i>t</i> _{1,3}	<i>t</i> _{2,3}	...	<i>t</i> _{<i>M</i>,3}
⋮	⋮	⋮	⋮	⋮	⋮
<i>x</i> _{<i>N</i>}	<i>y</i> _{<i>N</i>}	<i>t</i> _{1,<i>N</i>}	<i>t</i> _{2,<i>N</i>}	...	<i>t</i> _{<i>M</i>,<i>N</i>}

Keterangan:

X_{mn} = bobot TF-IDF pada kata kunci ke- n dan *tweet* ke- m

Langkah-langkah analisis data yang akan dilakukan dalam penelitian ini akan dijelaskan sebagai berikut.

1. Mengambil data *tweet* dengan Twitter API dan menyimpannya pada database
2. Menyiapkan data *tweet*, daftar *stopwords*, dan kata dasar.
3. Memberikan label awal sentimen data *tweets*.
4. Melakukan analisis *Text Preprocessing* yang terdiri dari *cleansing*, *case folding*, *tokenizing*, *filtering*, dan *stemming*.
5. Klasifikasi data menggunakan *Naive Bayes Classifier* untuk masing-masing instrumen.
6. Klasifikasi data menggunakan SVM kernel *linier*.
7. Klasifikasi data menggunakan SVM kernel RBF.
8. Membandingkan ketepatan klasifikasi NBC dan SVM.
9. Melakukan visualisasi dengan *word cloud*.
10. Menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

A. Preprocessing dan Karakteristik Data

Data *tweet* mengenai Saham GOTO Group (\$GOTO) pada media sosial *twitter* yang telah dikumpulkan dari Twitter API dilakukan *text preprocessing*. Berikut merupakan contoh salah satu *tweet* sebelum dilakukan *preprocessing text*.

Tabel 4 Data Tweet Sebelum Text Preprocessing

Tweet		Sentimen
GoTo Masuk Daftar Berpengaruh di Dunia	100	Positif
https://t.co/IKpvKen53P		

Pada Tabel 4 menampilkan contoh *tweet* awal yang didapatkan beserta sentimen dari *tweet* tersebut. Data tersebut kemudian dilakukan *text preprocessing* meliputi *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Berikut merupakan rangkaian *text preprocessing* yang dilakukan pada *tweet* ditampilkan pada Tabel 5

Tabel 5 Data *Tweet* Sebelum dan Sesudah *Text Preprocessing*

Tahapan	Sebelum	Sesudah
<i>Cleaning</i>	GoTo Masuk Daftar 100 Perusahaan Berpengaruh di Dunia https://t.co/lKpyKgn53P	GoTo Masuk Daftar Perusahaan Berpengaruh di Dunia
<i>Case Folding</i>	GoTo Masuk Daftar Perusahaan Berpengaruh di Dunia	goto masuk daftar perusahaan berpengaruh di dunia
<i>Tokenizing</i>	goto masuk daftar perusahaan berpengaruh di dunia	['goto', 'masuk', 'daftar', 'perusahaan', 'berpengaruh', 'di', 'dunia']
<i>Stopword</i>	['goto', 'masuk', 'daftar', 'perusahaan', 'berpengaruh', 'di', 'dunia']	['masuk', 'perusahaan', 'berpengaruh', 'dunia']
<i>Stemming</i>	['masuk', 'perusahaan', 'berpengaruh', 'dunia']	['masuk', 'usaha', 'pengaruh', 'dunia']

Pada Tabel 5 menunjukkan kalimat *tweet* yang telah dilakukan *preprocessing* hingga diperoleh hasil akhir berupa kata-kata yang terpisah. Berikut adalah struktur data *training* yang telah dilakukan *preprocessing teks* dengan menggunakan contoh *tweet*.

Tabel 6 Struktur Data Setelah *Text Preprocessing*

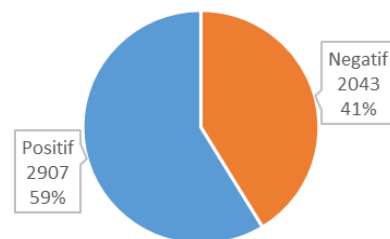
	besar	dunia	masuk	sukses	usaha
'masuk', 'usaha', 'pengaruh', 'dunia'	0	1	1	0	1
'oleh', 'nyata', 'efektif', 'masuk', 'besar', 'dunia'	1	1	1	0	0
...
'fakta', 'sukses', 'besar', 'tiga', 'asia'	1	0	0	1	0

Pada Tabel 6 menunjukkan struktur data yang telah berbentuk *document term matrix* berupa kata sejumlah 5846 kata. Kemudian pada kata-kata tersebut dilakukan pembobotan *Term Frequency Inverse Document Frequency* (TF-IDF) dengan hasil pembobotan pada Tabel 7

Tabel 7 Struktur Data TF-IDF

	besar	dunia	masuk	sukses	usaha
'masuk', 'usaha', 'pengaruh', 'dunia'	0	0,549	0,424	0	0,404
'oleh', 'nyata', 'efektif', 'masuk', 'besar', 'dunia'	0,345	0,376	0,291	0	0
...
'fakta', 'sukses', 'besar', 'tiga', 'asia'	0,381	0	0	0,426	0

Setelah langkah *text preprocessing*, selanjutnya dilakukan tahap analisis *tweet* terkait saham GOTO yang berupa sentimen positif dan negatif. Berikut merupakan sebaran label sentimen saham GOTO pada rentang waktu tanggal 14 Maret 2022 hingga 2 Mei 2022.



Gambar 4 Pie Chart Sentimen Tweets

Pada Gambar 4 menunjukkan sebaran *tweet* yang cenderung hampir seimbang dimana jumlah *tweet* bersentimen positif sedikit lebih banyak dibandingkan *tweet* bersentimen negatif. Jumlah *tweet* yang hampir seimbang ini menandakan keberagaman sentimen atau pendapat masyarakat terkait proses IPO GOTO. Dengan cenderung sebangnya jumlah antara sentimen positif dan negatif, maka data *tweet* ini dapat dikatakan *balance*. Keadaan *balance* pada kategori data tersebut akan berpengaruh pada perhitungan ketepatan klasifikasi. Ukuran ketepatan klasifikasi *accuracy* sesuai untuk data *balance* sehingga dapat dijadikan sebagai pengukuran ketepatan klasifikasi.

B. Klasifikasi Menggunakan NBC

Naïve Bayes Classifier (NBC) merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi yang kemudian dilakukan klasifikasi data pada kategori yang paling tepat. Data yang digunakan adalah data *tweet* yang sudah dilakukan pembobotan *term frequency-inverse frequency* (TF-IDF) sebagai variabel prediktor dan data sentimen *tweet* sebagai variabel respon. Variabel t merupakan sekumpulan dokumen bobot kata kunci yang direpresentasikan dengan pasangan atribut t_m untuk $m=1,2,...,M$. sedangkan y_k adalah himpunan kategori *tweet* dengan $k=-1$ adalah kategori negatif dan $k=1$ adalah kategori positif. Berikut merupakan hasil perhitungan nilai probabilitas sebuah *tweet* pada setiap kategori $P(Y_k)$ sesuai persamaan (4) pada Tabel 8 berikut

Tabel 8 Nilai $P(Y_k)$ Pada NBC

Sentimen	$P(Y_k)$
Positif	$P(Y_k=1) = 0,5845$
Negatif	$P(Y_k=-1) = 0,4155$

Berdasarkan Tabel 8 menunjukkan nilai probabilitas sebuah *tweet* bersentimen positif $P(Y_k=1)$ senilai 0,5845. Sedangkan nilai probabilitas *tweet* bersentimen negatif $P(Y_k=-1)$ senilai 0,4155. Kemudian untuk menentukan model NBC perlu menghitung probabilitas setiap kata pada masing-masing kategori $P(t_m/Y_k)$ sesuai persamaan (6) pada Tabel 9 berikut.

Tabel 9 Nilai $P(t_m/Y_k)$ Pada NBC

Sentimen	Kata Kunci				
	aamiin	ababil	...	zonk	zuper
Positif	0,000103	0,000068	...	0,000034	0,000034
Negatif	0,000042	0,000042	...	0,000208	0,000042

Berdasarkan Tabel 9 dapat diketahui nilai probabilitas setiap kata pada setiap kategori $P(t_m/Y_k)$. Model NBC menghasilkan nilai probabilitas yang digunakan untuk

mengklasifikasikan kategori *tweet* ke dalam sentimen positif atau negatif. Model NBC yang terbentuk berdasarkan pada Tabel 8 menunjukkan nilai $P(Y_k)$ sesuai persamaan (4) dengan Tabel 9 menunjukkan nilai $P(t_m|Y_k)$ sesuai persamaan (5). Berdasarkan model NBC didapatkan nilai Y_{MAP} sesuai persamaan (6) untuk masing-masing sentimen. Berikut adalah nilai Y_{MAP} dan kategori *tweet* ditunjukkan pada Tabel 10.

Tabel 10 Perhitungan Nilai Y_{MAP}

Data Ke-	$P(y_i = 1) \prod_{k=1}^n P(t_k y_i = 1)^{y_{ik}} P(y_i = 0) \prod_{k=1}^n P(t_k y_i = 0)^{1-y_{ik}}$	Y_{MAP}	Kategori	
0	7.71E-14	1.37E-13	7.71E-14	Negatif
2	4.13E-09	1.43E-09	4.13E-09	Positif
3	1.71E-09	6.17E-09	1.71E-09	Negatif
4	1.72E-19	2.10E-21	1.72E-19	Positif
⋮	⋮	⋮	⋮	⋮
4947	1.91E-22	5.18E-20	1.91E-22	Negatif
4948	9.73E-17	1.24E-18	9.73E-17	Positif
4949	4.14E-12	1.11E-10	4.14E-12	Negatif

Berdasarkan Tabel 10 didapatkan nilai probabilitas pada masing-masing kategori *tweet* kemudian didapatkan nilai Y_{MAP} dan prediksi kategori dari masing-masing *tweet*. Dari hasil prediksi tersebut akan dapat dibandingkan dengan kategori aktualnya sehingga nantinya akan diketahui besaran ketepatan klasifikasi model tersebut. Pembagian data menjadi data *training* dan data *testing* menggunakan metode *10-fold cross validation* dengan hasil sebagai berikut.

Tabel 11 Nilai Accuracy Metode NBC

Fold ke-	Training	Testing
1	91,65%	83,03%
2	91,67%	84,85%
3	92,01%	80,40%
4	91,83%	81,62%
5	91,38%	79,39%
6	91,81%	80,20%
7	91,36%	82,20%
8	91,56%	82,22%
9	91,66%	78,59%
10	91,65%	78,99%
Rata-rata	91,66%	81,15%

Dari Tabel 11 menunjukkan rata-rata nilai kebaikan klasifikasi NBC pada saham GOTO sebesar 91,66% pada data *training* dan 81,15% pada data *testing*, Nilai *accuracy* data *testing* tertinggi pada *fold* ke-2 dengan nilai *accuracy* sebesar 84,85%. Selanjutnya dilakukan pengukuran ketepatan klasifikasi menggunakan *confusion matrix* pada Tabel 12 berikut.

Tabel 12 Confusion Matrix Metode NBC Fold Ke-2

Data	Jumlah Tweet	True Positive	True Negative	False Positive	False Negative
Training	4455	2505	1582	107	264
Testing	495	269	151	29	46

Setelah didapatkan *confusion matrix*, langkah selanjutnya dilakukan perhitungan ketepatan klasifikasi *tweet* saham GOTO berdasarkan data *training* dan *testing*. Perhitungan ketepatan klasifikasinya disajikan pada Tabel 13 berikut.

Tabel 13 Ketepatan Klasifikasi NBC Fold Ke-2

Data	Accuracy	AUC
Training	91,67%	90,80%
Testing	84,85%	83,46%

Ukuran ketepatan klasifikasi pada data seimbang (*balance*) adalah *accuracy*. Berdasarkan Tabel 13 menunjukkan bahwa pada klasifikasi NBC data *training* diperoleh nilai *accuracy* sebesar 91,67% dan pada data *testing* diperoleh nilai *accuracy* sebesar 84,85%.

C. Klasifikasi Menggunakan SVM

Support Vector Machine (SVM) merupakan sistem pembelajaran yang bertujuan membangun pemisah optimum. SVM memisahkan data untuk mengklasifikasikan data. Pembahasan SVM ini meliputi Kernel *Linear* dengan parameter *C*, serta Kernel *Radial Basis Function* (RBF) dengan parameter *C* dan *gamma* (γ).

Pada SVM Kernel Linier akan menggunakan parameter *C* dengan nilai mulai dari 10^{-2} sampai dengan 10^3 . Data dibagi menjadi data *training* dan *testing* dengan menggunakan *10-fold cross validation*. Selanjutnya untuk penentuan parameter *C* optimal dilihat dari nilai ketepatan klasifikasi tertinggi.

Tabel 14 Penentuan Parameter SVM Kernel Linier

C	Ketepatan Klasifikasi
0,01	0,500
0,1	0,643
1	0,810
10	0,787
100	0,757
1000	0,749

Pada Tabel 14 didapatkan hasil klasifikasi optimum pada sentimen *tweet* GOTO Group dengan menggunakan SVM Kernel *Linear* pada parameter *C*=1. Selanjutnya dilakukan analisis kebaikan data pada setiap *fold* pembagian data *training* dan *testing* dengan menggunakan *k-fold validation* dengan hasil pada Tabel 15.

Tabel 15 Nilai Accuracy Metode SVM Kernel Linier

Fold ke-	Training	Testing
1	94,55%	84,04%
2	94,37%	83,64%
3	94,57%	82,42%
4	94,32%	83,23%
5	94,39%	80,61%
6	94,52%	81,82%
7	94,34%	84,65%
8	94,25%	82,63%
9	94,77%	79,19%
10	94,50%	80,40%
Rata-rata	94,46%	82,26%

Dari Tabel 15 menunjukkan rata-rata *accuracy* metode SVM Kernel Linier sebesar 82,26%. Nilai *accuracy* tertinggi pada *fold* ke-7 sebesar 84,65%. Selanjutnya dilakukan pengukuran ketepatan klasifikasi menggunakan *confusion matrix* yang ditampilkan pada Tabel 16 berikut.

Tabel 16 Confusion Matrix Metode SVM Kernel Linier Fold ke-7

Data	Jumlah Tweet	True Positive	True Negative	False Positive	False Negative
Training	4455	2499	1704	122	130
Testing	495	257	162	29	47

Setelah didapatkan *confusion matrix*, langkah selanjutnya dilakukan perhitungan ketepatan klasifikasi *tweet* saham GOTO berdasarkan data *training* dan *testing*. Perhitungan ketepatan klasifikasinya disajikan pada Tabel 17 berikut.

Tabel 17 Ketepatan Klasifikasi SVM Kernel Linier *Fold* ke-7

Data	Accuracy	AUC
Training	94,34%	94,13%
Testing	84,65%	83,69%

Ukuran ketepatan klasifikasi pada data seimbang (*balance*) adalah menggunakan *accuracy*. Berdasarkan Tabel 17 menunjukkan bahwa pada klasifikasi SVM Kernel Linier data *training* diperoleh nilai *accuracy* sebesar 94,34% dan pada data *testing* diperoleh nilai *accuracy* sebesar 84,65%. Sedangkan ada SVM Kernel RBF akan menggunakan parameter *C* dan *gamma* (γ) masing-masing mulai dari 10^{-2} sampai dengan 10^3 . Data dibagi menjadi data *training* dan *testing* dengan menggunakan *10-fold cross validation*. Selanjutnya penentuan parameter *C* dan γ optimal dilihat dari ketepatan klasifikasi tertinggi. Berikut merupakan ketepatan klasifikasi pada masing-masing nilai parameter yang disajikan pada Tabel 18.

Tabel 18 Penentuan Parameter SVM Kernel RBF

Parameter	<i>Gamma</i> (γ).					
	0,01	0,1	1	10	100	1000
<i>C</i>	0,01	0,500	0,500	0,500	0,500	0,500
	0,1	0,500	0,500	0,514	0,500	0,500
	1	0,500	0,722	0,813	0,531	0,530
	10	0,745	0,801	0,811	0,531	0,530
	100	0,800	0,785	0,811	0,531	0,530
	1000	0,776	0,784	0,811	0,531	0,530

Pada Tabel 18 didapatkan hasil klasifikasi optimum dengan menggunakan metode SVM Kernel RBF pada parameter *C*=1 dan γ =1. Selanjutnya dilakukan analisis kebaikan data dengan nilai *accuracy* pada setiap *fold* pembagian data *training* dan *testing* dengan menggunakan *k-fold validation*. Nilai kebaikan klasifikasi pada setiap *10-fold* disajikan pada Tabel 19 berikut.

Tabel 19 Nilai *Accuracy* Metode SVM Kernel RBF

<i>Fold</i> ke-	Training	Testing
1	98,56%	85,86%
2	98,61%	83,84%
3	98,61%	82,42%
4	98,45%	82,22%
5	98,63%	81,82%
6	98,63%	83,23%
7	98,50%	84,24%
8	98,61%	83,23%
9	98,63%	82,22%
10	98,70%	81,01%
Rata-rata	98,59%	83,01%

Dari Tabel 19 Menunjukkan rata-rata nilai *accuracy* metode SVM Kernel RBF pada saham GOTO sebesar 83,01%. Nilai *accuracy* tertinggi pada *fold* ke-1 sebesar 85,86%. Selanjutnya dilakukan pengukuran ketepatan klasifikasi menggunakan *confusion matrix* yang ditampilkan pada Tabel 20 berikut.

Tabel 20 *Confusion Matrix* Metode SVM Kernel EBF *Fold* ke-1

Data	Jumlah <i>Tweet</i>	True Positive	True Negative	False Positive	False Negative
Training	4455	4455	2579	1812	34
Testing	495	495	268	157	26

Setelah didapatkan *confusion matrix*, langkah selanjutnya dilakukan perhitungan ketepatan klasifikasi *tweet* saham GOTO berdasarkan data *training* dan *testing*. Perhitungan ketepatan klasifikasinya disajikan pada Tabel 21 berikut.

Tabel 21 Ketepatan Klasifikasi SVM Kernel RBF *Fold* ke-1

Data	Accuracy	AUC
Training	98,56%	98,54%
Testing	85,86%	84,63%

Ukuran ketepatan klasifikasi pada data seimbang (*balance*) adalah menggunakan *accuracy*. Berdasarkan Tabel 21 menunjukkan bahwa pada klasifikasi SVM Kernel RBF data *training* diperoleh nilai *accuracy* sebesar 98,56% dan pada data *testing* diperoleh nilai *accuracy* sebesar 85,86%.

Fungsi *hyperplane* pada SVM Kernel Linier didapatkan dengan mensubstitusikan fungsi kernel, sehingga didapatkan fungsi *hyperplane* untuk masing-masing kernel pada Tabel 22.

Tabel 22 Persamaan *Hyperplane*

Metode	Persamaan <i>Hyperplane</i>
SVM Kernel Linier	$f(t_k) = \sum_{m=1}^M \alpha_m y_m \left(\vec{t}_m \cdot \vec{t}_{m_2} \right) + 0,1091$
SVM Kernel RBF	$f(t_k) = \sum_{m=1}^M \alpha_m y_m \phi(t_{m_1}) \phi(t_{m_2}) - 0,3962$

Persamaan *hyperplane* pada Tabel 22, nilai α_m merupakan vector koefisien atau *lagrange multiplier* dari *support vector*. Nilai y_m merupakan label kategori yang memiliki dua nilai yaitu untuk positif dan negatif, serta t yang merupakan input yang akan diklasifikasikan. Sedangkan nilai 0,1091 pada SVM Kernel Linier dan nilai -0,3962 pada SVM Kernel RBF merupakan nilai bias pada masing-masing kernel.

D. Evaluasi Ketepatan Klasifikasi

Setelah mengetahui hasil masing-masing ketepatan klasifikasi pada metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM), maka langkah selanjutnya adalah membandingkan hasil dari kedua metode tersebut. Berikut merupakan hasil perbandingan metode NBC dan SVM berdasarkan nilai *accuracy* pada Tabel 23

Tabel 23 Perbandingan Ketepatan Klasifikasi NBC dan SVM

Data	<i>Accuracy</i>		
	NBC	SVM Linier	SVM RBF
Training	91,56%	94,37%	98,63%
Testing	84,44%	85,25%	85,86%

Tabel 23 menunjukkan nilai ketepatan klasifikasi terbaik adalah metode SVM Kernel RBF dengan nilai *accuracy* pada data *training* sebesar 98,63% dan pada data *testing* sebesar 85,86%. Hal ini menunjukkan bahwa pengklasifikasian sentimen positif dan negatif pada *tweet* mengenai saham GOTO Group sudah cukup baik.

E. Visualisasi Word Cloud

Visualisasi data teks menggunakan *word cloud* digunakan untuk mengetahui kata-kata yang paling sering muncul pada data. Berikut adalah *word cloud* mengenai sentimen positif masyarakat Indonesia terhadap Saham GOTO Group.



Gambar 5 *Word Cloud* Sentimen Positif

Gambar 5 menunjukkan kata-kata yang sering muncul pada sentimen positif yaitu kata “beli”, “gojek”, “dapat”, “harga”, “investor”, “tokopedia”, “mitra”, “indonesia”, “driver”, dan “bursa”. Kata-kata tersebut menunjukkan bahwa pengguna Twitter banyak membahas mengenai pencapaian perusahaan Gojek dan Tokopedia sebagai *startup* Indonesia yang berhasil melantai di bursa saham. Hal ini memunculkan antusiasme di kalangan investor untuk memesan, membeli, dan memantau pergerakan saham ini. Sedangkan *word cloud* mengenai sentimen negatif terhadap dalah sbegai berikut.



Gambar 6 *Word Cloud* Sentimen Negatif

Gambar 6 menunjukkan kata-kata yang sering muncul pada sentimen positif yaitu kata “bukalapak”, “harga”, “beli”, “rugi”, “turun”, dan “arb”. Kata-kata tersebut menunjukkan bahwa pengguna Twitter banyak membahas kekecewaan mereka terhadap penurunan harga saham GOTO yang menyebabkan kerugian bagi beberapa investor. Kejadian ini sering kali dibanding-bandingkan dengan kejadian *startup* serupa, Bukalapak, yang mengalami penurunan harga secara signifikan selepas *listing* di bursa saham.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil dan pembahasan diperoleh kesimpulan yaitu *tweet* tentang GOTO yang diunggah pada tanggal 14 Maret 2022 dan 2 Mei 2022 berjumlah 4950 *tweets* dengan sebaran *tweet* bersentiment positif dan negatif cukup berimbang. Setelah tahap *text preprocessing* didapatkan 5846 kata yang nantinya akan menjadi variabel prediktor. Hasil ketepatan klasifikasi menggunakan NBC pada data *testing* diperoleh rata-rata nilai *accuracy* sebesar 81,15%. Hasil ketepatan klasifikasi menggunakan SVM Kernel Linier dengan parameter optimal ($C=1$) diperoleh rata-rata nilai *accuracy* sebesar 82,26%. Sedangkan hasil ketepatan klasifikasi menggunakan SVM Kernel RBF dengan parameter optimal ($C=1$ dan $\text{Gamma}=1$) diperoleh nilai *accuracy* sebesar 83,01%. Perbandingan Ketepatan Klasifikasi NBC dan SVM menunjukkan hasil bahwa performa metode SVM Kernel RBF lebih baik dalam mengklasifikasikan data teks sentimen masyarakat terhadap Saham GOTO Group. Visualisasi *word*

cloud pada *tweet* bersentimen positif menunjukkan antusiasme atas pencapaian perusahaan Gojek dan Tokopedia sebagai *startup* Indonesia yang berhasil melantai di bursa saham. Sedangkan visualisasi *word cloud* pada *tweet* bersentimen negatif menunjukkan kekecewaan yang terjadi akibat penurunan harga saham GOTO yang serupa dengan kerugian yang dialami pada *startup* lain yaitu Bukalapak.

B. Saran

Saran kepada perusahaan GOTO Group untuk tetap menjaga *image* dan *brand* perusahaan di media sosial untuk terus mempertahankan dan meningkatkan kepercayaan investor sehingga harga saham bisa tetap terjaga dengan baik. Kemudian saran kepada penelitian berikutnya untuk menambahkan tahap *text preprocessing* lebih detail terutama mengatasi *slang word* dengan harapan dapat meningkatkan kebaikan dalam tahap analisis klasifikasi nantinya.

DAFTAR PUSTAKA

- [1] BEL. (2022). *Saham*. Retrieved from PT Bursa Efek Indonesia: <https://www.idx.co.id/produk/saham/>
- [2] Mandiri Sekuritas. (2022). *IPO Saham*. Retrieved from MOST: <https://www.most.co.id/bantuan/ipo-saham>
- [3] GoTo. (2022). *Tentang Kami*. Retrieved from GOTO: <https://www.gotocompany.com/about-us>
- [4] Buntoro, G. A., Adji, T. A., & Purnamasari, A. E. (2014). Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. *Conference on Information Technology and Electrical Engineering*, 38-43.
- [5] Twitter. (2022). *About Twitter*. Retrieved from Twitter: <https://about.twitter.com/en>
- [6] Liu. (2012). *Sentiment Analysis and Opinion Mining*. USA: Morgan & Claypool.
- [7] Dehaff, M. (2010). *Sentiment Analysis, Hard But Worth It!* USA: Cambridge University.
- [8] Ian, H. W., Eibe, F., & Mark, A. H. (2011). *Data Mining : Practical Machine Learning Tools and Techniques*. USA: Morgan Kauffmann.
- [9] Asian, J. (2007). *Effective Techniques for Indonesian Text Retrieval*. Phd Thesis of School of Computer Science and Information Technology RMIT University Australia
- [10] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann.
- [11] Spiegelhalter, D., & Rice, K. (2006). *Bayesian Statistics. Scholarpedia*, 4(8), 5230.
- [12] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann.
- [13] Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- [14] Bekkar, M., Djemma, H. K., & Alitouche, T. A. (2013). Evaluation Measure for Models Assesment over Imbalanced Data Sets. *Journal of Information Engineering and Applications* 3, 27-38.
- [15] Williams, G. (2011). *Data Mining with Ratle and R : The Art of Excavating Data for Knowledge Discovery*. New York: Springer.
- [16] Twitter. (2022). *About the Twitter API*. Retrieved from Twitter: <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api#Twitter>
- [17] Gojek. (2022). *Gojek and Tokopedia combine to form GoTo, the largest technology group in Indonesia and the "go to" ecosystem for daily life*. Retrieved from Gojek: <https://www.gojek.com/blog/gojek/goto/>