

## ANALISIS SENTIMEN PENGGUNA GOPAY MENGGUNAKAN METODE LEXICON BASED DAN SUPPORT VECTOR MACHINE

**Rachmad Mahendrajaya\*, Ghulam Asrofi Buntoro, Moh. Bhanu Setyawan**

Fakultas Teknik, Universitas Muhammadiyah Ponorogo  
E-mail Korespondensi : rachmadmj@gmail.com

History Artikel

Diterima: 27 Agustus 2019    Disetujui: 17 September 2019    Dipublikasikan: 07 Oktober 2019

---

### *Abstract*

*Go-Pay is part of the Gojek application and one of the most popular finteches in Indonesia. Although the most popular, not all users have positive or even negative comments. Now users can submit various media opinions, one of which is Twitter. Twitter media has the advantage of a simple display, updated topics, open access to tweets and express opinions quickly. From a variety of comments on Twitter it takes a technique to divide into classes positive or negative opinions. This study uses preprocessing and labeling opinions into positive and negative classes with the lexicon Based method. As for the classification using the Support Vector Machine (SVM) method. The data used in the form of opinions about Go-Pay reviews from social media Twitter, amounting to 1210. The results of labeling with Lexicon Based amounted to 923 for positive and 287 for negative. While the classification of the SVM method using the Linear kernel produces 89.17% and 84.38% for the Polynomial kernel.*

**Keywords:** *Sentiment Analysis, Twitter, Go-Pay, Lexicon Based, Support Vector Machine (SVM)*

#### Abstrak

Go-Pay merupakan bagian dari aplikasi Gojek dan salah satu *fintech* yang paling diminati di Indonesia. Meskipun paling diminati, tidak semua pengguna berkomentar positif bahkan negatif. Kini pengguna dapat menyampaikan opini berbagai media salah satunya Twitter. Media twitter memiliki kelebihan tampilan simple, topik terupdate, terbuka mengakses tweet dan cepat dalam menyampaikan opini. Dari berbagai ragam komentar di Twitter diperlukan teknik untuk membagi ke dalam kelas opini positif ataupun negatif. Penelitian ini menggunakan preprocessing dan melabeli opini ke dalam kelas positif dan negatif dengan metode *lexicon Based*. Sedangkan untuk klasifikasinya menggunakan metode *Support Vector Machine* (SVM). Data yang digunakan berupa opini tentang ulasan Go-Pay dari media sosial Twitter yang berjumlah 1210. Hasil dari pelabelan dengan *Lexicon Based* berjumlah 923 untuk positif dan 287 untuk negatif. Sedangkan klasifikasi metode SVM menggunakan kernel Linear menghasilkan 89,17% dan 84,38% untuk kernel Polynomial.

**Kata Kunci :** *Analisis Sentimen, Twitter, Go-Pay, Lexicon Based, Support Vector Machine (SVM)*

**Mahendrajaya, Rachmad** (2019). *Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine*. KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo, 3(2), 2019: 52-63

© 2019 Universitas Muhammadiyah Ponorogo. All rights reserved

---

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

## A. PENDAHULUAN

Perkembangan Teknologi komunikasi saat ini berkembang semakin pesat dan cepat ke arah serba digital. Meningkatnya kebutuhan informasi dan data mendorong manusia untuk membuat teknologi baru agar pengolahan data dan informasi dapat dilakukan dengan mudah dan cepat (Josi,2014).

Dari hasil survei yang diselenggarakan oleh Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) ditahun 2017 dan 2018 terakhir didapatkan pengguna internet di bagi beberapa bidang. Salah satunya transaksi uang secara online mencapai 47,7% dari pengguna internet Indonesia. Dimana transaksi online ini cukup banyak penggunaannya salah satunya memanfaatkan dompet digital sebagai media transaksi.

Dengan layanan dompet digital, pengguna dapat melakukan transaksi secara nontunai tanpa kartu baik online maupun offline. Diindonesia sendiri banyak layanan dompet digital atau fintech sebut saja Go-Pay, OVO, T-cash, Dana dan sebagainya. Dari survei yang dilakukan DailySocial pada tahun 2018 menandakan Go-Pay banyak diminati oleh masyarakat Indonesia, karena prosentase yang dihasilkan lebih tinggi dari pada yang lainnya.

Go-Pay sendiri masih satu layanan dengan Go-Jek yang didirikan oleh Nadiem Makarim pada tahun 2010. Selama beroperasi, Gojek dikenal sebagai perusahaan transportasi *online* yang cukup besar di Indonesia. Setelah bisnis transportasi yang cukup berhasil, kini Gojek mulai merambah industri fintech dengan pengembangan fungsi Go-Pay. Pada tahun 2016 aplikasi Gojek resmi meluncurkan layanan Go-Pay. Setelah mendapat izin dari Bank Indonesia sebagai penyedia uang elektronik, kini saldo Go-Pay dapat digunakan untuk alat pembayaran didalam aplikasi Go-Jek, toko online maupun offline yang telah bekerjasama dengan pihak Gojek. Meskipun dengan kemudahan yang diberikan oleh layanan Go-Pay, tentu tidak semua memberikan komentar positif terhadap layanan Go-Pay, mungkin beberapa ada yang memberikan komentar negatif bahkan

netral. Masyarakat kini dapat memberikan komentarnya di berbagai media. Salah satu media yang banyak digunakan untuk menyampaikan opini adalah media sosial khususnya *Twitter*.

Twitter merupakan salah satu media sosial yang memiliki pengguna perharinya 126 juta. Selain itu twitter memiliki Twitter Trending Topic yang merupakan daftar topik terpopuler dan diupdate setiap waktu. Selanjutnya kelebihan lain dari twitter semua postingan yang ada di Twitter dapat diakses oleh pengguna lainnya. Dengan kemudahan tersebut pengguna Twitter dapat saling berinteraksi tanpa harus menjadi follower. Kelebihan lainnya dari sifat terbuka Twitter yaitu dapat membuat tweet dengan menandai (mention) siapapun tanpa harus membuat tweet di fans page atau akun pihak yang ditandai tersebut.

Dari pemaparan sebelumnya twitter memiliki kelebihan dibanding media sosial lainnya yaitu tampilan yang simple, topik terupdate, cepat dalam mengabarkan sebuah opini pengguna dan terbuka untuk mengakses tweet. Tentu ini semakin memudahkan dalam pengumpulan data opini dari masyarakat dan sebagai media penelitian. Agar dapat mengetahui kepuasan pengguna terhadap layanan Go-Pay diperlukan analisis sentimen terhadap twitter. Analisis sentimen dapat mengelompokkan opini pada kalimat untuk mengetahui sentimen positif atau negatif. Untuk analisis sentimen itu sendiri dapat menggunakan beberapa metode salah satunya metode *Lexicon Based* dan *Support Vector Machine (SVM)*.

Tujuan dari penelitian adalah sebagai berikut :

1. Menghasilkan data analisis sentimen berupa komentar pengguna terhadap layanan Go-Pay melalui Twitter dengan memanfaatkan metode *Lexicon Based* untuk menentukan class sentiment dengan jumlah data yang diambil 1210 dan dibatasi pada 24 juli 2019 sampai 30 juli 2019.
2. Pengambilan data dari Twitter dan analisa data menggunakan R Studio.

3. Mengetahui tingkat akurasi 2 kernel dari metode *Support Vector Machine (SVM)* untuk klasifikasi komentar pengguna Go-Pay.

## B. KAJIAN LITERATUR

### 1. Perkembangan Dompot Digital Go-Pay

Kemajuan teknologi membuat segala macam layanan diakses secara digital. Mulai dari transportasi hingga transaksi mengalami digitalisasi. Saat ini uang dalam bentuk fisik mulai dikurangi penggunaannya karena kurang efektif jika dibawa dalam jumlah banyak. Selaras dengan perkembangan dompet digital di Indonesia yang mulai diminati masyarakat salah satunya adalah Go-Pay. Go-Pay merupakan layanan dompet digital yang dimiliki perusahaan Gojek. Pada awalnya Go-Pay hanya bisa digunakan untuk pembayaran jasa transportasi saja. Namun Nadiem CEO Gojek ingin Gojek lebih berkembang dari sekedar transportasi online, Gojek ingin bertransformasi menjadi sebuah fintech atau financial technology melalui layanan Go-Pay. Go-Pay nantinya bisa digunakan sebagai alat pembayaran untuk transaksi di luar layanan aplikasi seperti ditunjukkan pada Gambar 1 ([www.tirto.id](http://www.tirto.id)).



Gambar 1 Evolusi Go-Pay

### 2. Lexicon Based

Lexicon Based merupakan metode umum untuk analisis sentimen pada media sosial dikarenakan metode yang digunakan praktis. Metode yang digunakan Lexicon Based

menggunakan kamus sebagai sumber bahasa atau leksikal. Lexicon Based sebagai pengklasifikasi sebuah sentimen dari setiap opini sehingga sebuah kalimat sentimen dapat di kelas-kelaskan sesuai kelas negatif dan positif. Dengan metode ini sebuah data kalimat sentimen akan lebih mudah untuk di klasifikasi.

### 3. SVM

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Menurut (Buntoro, 2017) *Support Vector Machine (SVM)* adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. Algoritma SVM asli diciptakan oleh Vladimir Vapnik dan turunan standar saat ini (margin lunak) diusulkan oleh Corinna Cortes dan Vapnik Vladimir (Buntoro, 2017).

## C. PELAKSANAAN DAN METODE

### 1. Pengumpulan Data

Pada tahap awal pengumpulan data tweet, data diambil dengan metode crawling dari media sosial Twitter menggunakan bantuan R Studio. Sebelum proses crawling diperlukan API Key dari Twitter sebagai autentifikasi dengan R Studio. Setelah mendapat API Key selanjutnya diperlukan package untuk crawling twitter di R Studio. Package diinstal dan proses crawling data dari twitter dapat dijalankan.

### 2. Proses Pengumpulan Data

Setelah mendapat API Key twitter selanjutnya kita perlu menginstal package yang dibutuhkan R Studio yaitu httr package dan rewet package.

### 3. Preprocessing Data

Preprocessing data tweet yang sebelumnya dari data tidak terstruktur menjadi data yang terstruktur sehingga data dapat dianalisis.

#### 4. Kelas Opini

Setelah data melewati tahap preprocessing kemudian data tweet akan ditentukan kelas opini yang dimunculkan dalam penelitian ini ada dua yaitu positif dan negatif. Proses pelabelan data dilakukan oleh *kamus lexicon based* dengan cara menghitung skor sentimen. Dengan 2 kelas opini mampu memberi penilaian layanan perusahaan terhadap kepuasan masyarakat.

#### 5. Load Dictionary

Pada proses load dictionary membutuhkan kamus untuk proses pembobotan. Banyak kamus bisa digunakan namun pada penelitian ini menggunakan salah satu kamus opini positif dan negatif keyword dari penelitian (Liu, Bing, Hu, Minqing, and Cheng, Junsheng (2005), yang telah diterjemahkan ke dalam bahasa Indonesia.

#### 6. Pembobotan Lexicon Based

*Lexicon Based* merupakan fitur kata yang memiliki sentimen positif atau negatif berdasarkan kamus atau *lexicon*. Proses pelabelan data dilakukan oleh *kamus lexicon based* dengan cara menghitung skor sentimen. Setelah diketahui mana kata yang memuat positif, negatif dan netral di dalam sebuah kalimat, selanjutnya menghitung setiap kata yang mengandung sentimen dalam kalimat tersebut, dengan menjumlahkan nilai opini. Jumlah nilai opini untuk sentimen positif bernilai 1 atau lebih, sedangkan netral dalam kalimat tersebut untuk yang bernilai = 0, begitupun sebaliknya untuk negatif dalam kalimat tersebut bernilai = -1 atau lebih (Buntoro, 2017).

#### 7. Klasifikasi

Setelah pembobotan menggunakan metode lexicon based sekarang masuk proses klasifikasi untuk menghitung akurasi dengan metode Suport Vector Machine. Pada tahap ini menggunakan bantuan Weka 3.8.2 untuk menghitung

akurasi metode SVM. SVM sendiri pada pemaparan sebelumnya adalah mencari *hyperplane* pemisah dua kelas data yang terbaik.

Dengan persamaan Suport Vector Machine sendiri disini adalah

$$f(x) = w \cdot x + b \quad \text{atau} \quad f(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b$$

Keterangan :

- $w$  : Parameter *hyperplane* yang dicari (garis yang tegak lurus antara garis *hyperplane* dan titik *support vector*)
- $x$  : Titik data masukan Support Vector Machine
- $\alpha_i$  : Nilai bobot setiap titik data
- $(x, x_i)$  : Fungsi kernel
- $b$  : parameter *hyperplane* yang dicari (nilai bias)

#### 8. Evaluasi Model Klasifikasi

Pengukuran terhadap kinerja sistem klasifikasi perlu dilakukan. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data.

### D. HASIL DAN PEMBAHASAN

#### 1. Pengumpulan Data

Pada tahap pengumpulan data twitter proses yang perlu dilakukan terlebih dahulu adalah mendapatkan API Key Twitter yang bisa didapatkan di <https://developer.twitter.com/apps>. Tahap pertama membuat aplikasi pada App Twitter, lalu mengisi form pengajuan API Key. Setelah berhasil mendaftar maka akan mendapatkan API Key yang berguna menghubungkan R Studio dengan Twitter.

Setelah mendapat API Key proses crawling dapat dilakukan. Pertama memanggil library di R Studio yang sudah diinstal yaitu "httr" dan "rtweet". Apabila library sudah diaktifkan, selanjutnya proses autentifikasi API Key yang telah didapat sebelumnya.

Setelah autentifikasi API Key berhasil, proses crawling dapat dilakukan

dengan mencari kata “gopay”, data yang terkumpul berjumlah 1210 dengan bahasa Indonesia. Data diambil dari tanggal 24 Juli 2019 sampai 30 Juli 2019. Hasil crawling dapat dilihat pada Gambar 2.

created_at	screen_name	text
1 2019-07-30 23:56:10	devillaaa	@ohmybeautybank Payment pake gopay masih cashbac...
2 2019-07-30 23:54:09	MSNIndonesia	Gopay Kini Bisa Jadi Alat Pembayaran di Google Play htt...
3 2019-07-30 23:47:19	moanaandthe	pengen GA skin care <U-274E> pengen GA Pulsa Ovo ...
4 2019-07-30 23:42:11	Y_Vendy	@asianweirdos Gopay Semoga kuliah &amp rezekimu la...
5 2019-07-30 23:41:25	titipdongkak	@asianweirdos Gopay hehe semoga lancar kuliahnya ya...
6 2019-07-30 23:37:57	Redscholars	@rencez_23 Gw bayar cash juga itu gak pake gopay. Nya...
7 2019-07-30 23:37:39	amaliaapratiwi	lama2 bingung sama gopay isi 30rb yg masuk 29rb. Ada ...
8 2019-07-30 23:36:18	hellyeih	@imismeeeh @budakkafein bismillah gopay
9 2019-07-30 23:31:44	araachii	@asianweirdos selamat!! Semoga kuliah dan rezekinya ...
10 2019-07-30 23:27:05	weirdaq	@matchulatte @dicriey gopay 50k butuh bgt buat go ri...
11 2019-07-30 23:26:57	NoRini	@matchulatte gopay / pulsa. semangat terus ya kamu<...
12 2019-07-30 23:25:38	cikidaaw	yampun kaget pagiz dikirimin saldo gopay sama ortu m...

Gambar 2 Hasil Crawling

Selanjutnya data crawling disimpan di excel dengan format CSV perintah yang digunakan *save\_as\_csv()*. Untuk tahap selanjutnya data crawling dapat digunakan ke tahap Preprocessing.

## 2. Preprocessing

Pada tahap *Preprocessing*, data hasil crawling sebelumnya yang telah didapat masih data mentah dimana data tersebut masih banyak mengandung *noise* dan tidak terstruktur. Maka dari itu diperlukan *preprocessing* untuk menghilangkan *noise* dalam kata, mengkonversikan data agar lebih terstruktur dan untuk mempermudah penelitian. Pada penelitian, tahap *preprocessing* yang digunakan diantaranya normalisasi kalimat, *Cleansing*, menghilangkan angka, emoticon dan tanda baca, *case folding*, *filtering* dan *tokenizing*. Pada *preprocessing* ini dibantu dengan R Studio dan microsoft excel. Hasil *preprocessing* pada penelitian yang telah dilakukan seperti berikut.

### a. Normalisasi kalimat

Tahap awal dalam *preprocessing* dalam penelitian ini adalah normalisasi kalimat atau memperbaiki kata ejaan yang disingkat agar sesuai dengan KBBI dan mempermudah dalam penelitian.

Proses penerapan normalisasi kalimat seperti pada Tabel 1.

Tabel 1 Proses Normalisasi Kalimat

Input	Output
Dengan senang hati dan seikhlasnya trf ke gopay saya :) brp aja bebas <a href="https://t.co/46CaAEhrXO">https://t.co/46CaAEhrXO</a>	Dengan senang hati dan seikhlasnya transfer ke gopay saya :) berapa aja bebas <a href="https://t.co/46CaAEhrXO">https://t.co/46CaAEhrXO</a>
RT@_hellside Itu mah karena anda ga pernah ngisi gopay anda yah anak setan :)	RT@_hellside Itu mah karena anda tidak pernah ngisi gopay anda yah anak setan :)

### b. *Cleansing*

*Cleansing* merupakan membersihkan data tweet dari *noise* yang berupa mention username (@username), HTML, kata kunci, RT atau Retweet, hashtag (#), email, link tautan (<http://website.com>), dan mention (@username). *Noise* tersebut perlu dihilangkan agar mempermudah penelitian seperti ditunjukkan pada Tabel 2.

Tabel 2 Proses *Cleansing*

Input	Output
Dengan senang hati dan seikhlasnya transfer ke gopay saya :) berapa aja bebas <a href="https://t.co/46CaAEhrXO">https://t.co/46CaAEhrXO</a>	Dengan senang hati dan seikhlasnya transfer ke gopay :) saya berapa aja bebas
RT@_hellside Itu mah karena anda tidak pernah ngisi gopay anda yah anak setan :)	Itu mah karena anda tidak pernah ngisi gopay anda yah anak setan :)

### c. Menghilangkan angka, emoticon dan tanda baca

Menghilangkan angka, emoticon dan tanda baca perlu dilakukan dalam sentimen analisis. Dalam penelitian ini yang diperlukan berupa kata atau kalimat yang bersih, apabila ada angka, emoticon dan tanda baca maka dapat mengubah makna dalam kalimat tweet tersebut. Proses tersebut ditunjukkan Tabel 3.

Tabel 3 Proses menghilangkan angka, emoticon dan tanda baca

Input	Output
Dengan senang hati dan seikhlasnya transfer ke gopay saya :) berapa aja bebas	Dengan senang hati dan seikhlasnya transfer ke gopay saya berapa aja bebas
Itu mah karena anda tidak pernah ngisi gopay anda yah anak setan :)	Itu mah karena anda tidak pernah ngisi gopay anda yah anak setan

d. Case folding

Case folding merupakan proses menyeragamkan huruf kapital ke dalam huruf standar atau kecil semua. Dikarenakan penggunaan huruf kapital banyak tidak konsisten dan peletakannya. Proses case folding ditunjukkan Tabel 5.

Tabel 4. Proses case folding

Input	Output
Dengan senang hati dan seikhlasnya transfer ke gopay saya berapa aja bebas	dengan senang hati dan seikhlasnya transfer ke gopay saya berapa aja bebas
Itu mah karena anda tidak pernah ngisi gopay anda yah anak setan	itu mah karena anda tidak pernah ngisi gopay anda yah anak setan

e. Filtering

Tahap Filtering merupakan tahap penyaringan kalimat tweet atau pengurangan kata yang didalam *corpus* yang disebut *stopword*. *Stopword* sendiri adalah menghilangkan kata yang tidak berpengaruh atau tidak memiliki makna dalam penelitian. Dalam penelitian ini juga menghilangkan kata kunci dari proses crawling dan kata yang tidak diinginkan seperti gopay, ovo, dana dan Linkaja dengan menggunakan perintah *stopword*. Proses filtering seperti pada Tabel 5.

Tabel 5 Proses filtering

Input	Output
dengan senang hati dan seikhlasnya	senang hati seikhlasnya transfer berapa aja bebas

transfer ke gopay saya berapa aja bebas	
itu mah karena anda tidak pernah ngisi gopay anda yah anak setan	mah pernah ngisi yah anak setan

f. Tokenizing

Tahap terakhir dalam *preprocessing* ini adalah tokenizing. Tokenizing sendiri adalah proses memecah kata per kata pada sebuah kalimat atau dokumen. Proses ini diperlukan karena untuk mendapatkan kata yang memiliki nilai. Tokenizing ini dapat mempermudah menghitung frekuensi kemunculan kata dalam *corpus*. Proses tokenizing ditunjukkan pada Tabel 6.

Tabel 6 Proses tokenisasi

Input	Output
senang hati seikhlasnya transfer berapa aja bebas	senang, hati, seikhlasnya, transfer, berapa, aja, bebas,
mah pernah ngisi yah anak setan	mah, pernah, ngisi, yah, anak, setan,

Setelah *preprocessing* selesai maka tahap selanjutnya adalah membangun Term Document Matrix (TDM), TDM sendiri adalah bentuk matriks matematika yang menggambarkan frekuensi kata yang sering muncul dalam *corpus*, dalam penelitian ini menampilkan frekuensi kata dari 1210 tweet menghasilkan 3047 kata seperti ditunjukkan pada Gambar 3.

	word	freq
	<b>pakai</b>	245
	<b>aja</b>	175
	<b>bayar</b>	121
	<b>saldo</b>	117
	<b>kalo</b>	100
	<b>cashback</b>	86
	<b>isi</b>	74
	<b>gojek</b>	73
	<b>beli</b>	69
	<b>semoga</b>	67
	<b>via</b>	66
	<b>kak</b>	66

Showing 1 to 12 of 3,047 entries

Gambar 3 Frekuensi TDM

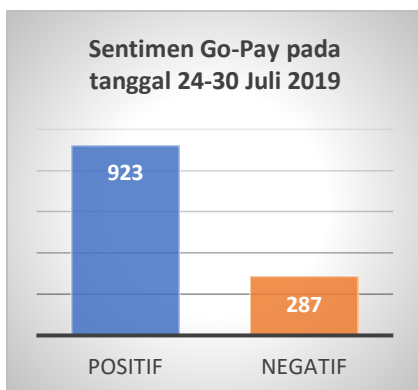
### 3. Pelabelan Kelas Opini

Pelabelan kelas opini atau sentimen dimana memberikan label pada kalimat atau dokumen ke dalam kelas positif atau negatif. Dalam penelitian ini menggunakan algoritma *lexicon based* dengan cara menghitung skor sentimen positif dikurangi skor sentimen negatif didalam kalimat. Hasil kalimat dari perhitungan skor yang  $\geq 0$  akan dilabeli kelas positif, begitupun sebaliknya jika hasilnya  $< 0$  maka dilabeli negatif. Dalam pelabelan kelas opini ini yang dibantu R Studio dengan menggunakan kamus positif dan negatif. Berikut hasil dari pelabelan yang diperoleh dari R ditunjukkan pada Gambar 4 .

score	text	klasifikasi
1	payment pakai cashback sis	Positif
2	-1 alat pembayaran google play	Negatif
3	0 pengen skin care ue pengen pulsa u dasar akuc	Positif
4	1 semoga kuliah rezekimu lancar amin	Positif
5	-2 hehe semoga lancar kuliahnya ya jurusan beda univ nih ...	Negatif
6	1 bayar cash gak pakai nyari point doang	Positif
7	-2 bingung isi rb masuk rb saldo langsung gimana sih	Negatif
8	1 bismillah	Positif
9	1 selamaat semoga kuliah rezekinya dilancarkan aamiin	Positif
10	1 k butuh banget go ride gak bawa kendaraan plus gaad...	Positif
11	0 pulsa semangat ya kamuuuu	Positif
12	0 yampun kaget pagi dikirimin saldo ortu murid d	Positif

Gambar 4. Hasil pelabelan dengan R

Setelah data hasil pelabelan didapat, selanjutnya semua data sentimen dijumlah sesuai dengan kelasnya. Hasil penjumlahan ditunjukkan pada diagram Gambar 5.



Gambar 5 Diagram jumlah data pelabelan Hasil pelabelan diatas menunjukkan dari data tweet berjumlah 1210, menghasilkan frekuensi sentimen positif berjumlah 923 yang lebih tinggi

daripada sentimen negatif yang berjumlah 287 ulasan. Dari pelabelan sentimen sebelumnya, bisa diambil kesimpulan bahwa layanan Go-pay masih tergolong baik atau positif. Meskipun masih banyak positifnya, tidak menutup kemungkinan untuk meningkatkan layanan untuk mengurangi sentimen negatif. Berikut beberapa tweet hasil pelabelan ditunjukkan pada Tabel 7 .

Tabel 7 Pelabelan sentimen tweet

tweet	sentimen	skor
rp gais pls promo gojek dongggg promo gojek nyaaaa	Positif	2
salah kirim isi ulang nomor hangus	Negatif	-2

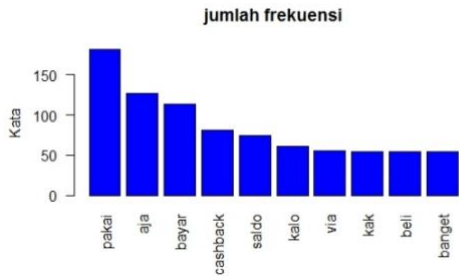
### 4. WordCloud Sentimen

Hasil dari pelabelan kelas sentimen didapati opini dari dua kelas positif dan negatif, namun kata opini ulasan tentang Go-Pay masih menjadi satu yang belum dipisahkan ke dalam kelas sentimen tersendiri sehingga kata yang sering muncul tersebut belum diketahui untuk ulasan positif maupun negatif. Maka itu diperlukan pembagian ulasan masing-masing ke dalam dua kelas sentimen. Untuk mengetahui kata yang sering muncul termasuk ke dalam kelas positif maupun negatif. Penelitian ini menggunakan visual *WordCloud* dari kedua sentimen yaitu kelas positif dan negatif.

#### a. Sentimen Positif

Dari sentimen positif yang didapatkan berjumlah 923 tweet di pecah menjadi perkata sehingga memunculkan frekuensi kata yang sering muncul dalam ulasan Go-Pay. Dari hasil beberapa kata dalam 10 besar untuk ulasan positif yang sering muncul ditampilkan kedalam diagram yang ditunjukkan pada Gambar 6.





Gambar 6 Frekuensi kata positif

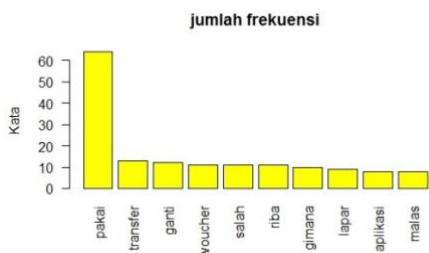
Selanjutnya dari kumpulan kata yang sering muncul dibangun menjadi *wordcloud* yang ditunjukkan pada Gambar 7.



Gambar.7 WordCloud positif

b. Sentimen Negatif

Sedangkan untuk sentimen negatif yang didapatkan 287 tweet dipecah menjadi beberapa kata yang ditampilkan ke dalam 10 besar untuk ulasan negatif dan ditampilkan ke bentuk diagram pada Gambar 8.



Gambar 8 Frekuensi kata negative

Setelah frekuensi muncul, selanjutnya membuat *wordcloud* dari beberapa kumpulan kata negatif yang ditunjukkan pada Gambar 9.



Gambar 9 WordCloud negatif

5. Klasifikasi Model SVM

Setelah pelabelan maka data tweet yang sudah dibagi kelasnya masuk ketahap klasifikasi dengan menghitung akurasi algoritma SVM. Didalam SVM terdapat berbagai kernel, sehingga penelitian ini membandingkan mana kernel yang memiliki akurasi yang cukup baik. SVM sendiri dapat mengklasifikasi data berdimensi tinggi, maka dari itu dalam penelitian ini proses klasifikasi dibantu dengan aplikasi WEKA. Sebelum proses klasifikasi, data dari pelabelan sebelumnya yang dihasilkan diubah menjadi vector, dalam weka terdapat tool yang mengubah kata string menjadi vector yaitu *StringToWordVector*.

Twitter-weka.filters.unsupervised.attribute.StringToW				
Numeric	Numeric	Numeric	Numeric	N
1659: pagi	1660: paha	1661: paham	1662: pakai	166
0.0	0.0	0.0	1.213887	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	1.213887	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
3.571417	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	
0.0	0.0	0.0	1.213887	
0.0	0.0	0.0	1.213887	

Gambar 10 Data Vector dari data Tweet

Gambar 10 menunjukkan bahwa kolom dengan kata “pakai” memiliki nilai 1.213887 pada baris ke-1 sedangkan yang lain bernilai 0,0 ini menunjukkan kata pakai terletak pada tweet ke-1, ke-6, ke-14, ke-15.

a. Pembobotan Kata

Dalam proses pembobotan kata (term) perlu dilakukan untuk klasifikasi, pembobotan dalam penelitian ini menggunakan TF-IDF (*Term Frequency – Inverse Document Frequency*). **TF (Term Frequency)** adalah frekuensi kemunculan sebuah term dalam dokumen yang bersangkutan sedangkan IDF adalah hubungan ketersediaan sebuah term dalam seluruh dokumen.

b. Pengujian Kernel SVM

Kernel Linier digunakan ketika data yang akan diklasifikasi dapat terpisah dengan sebuah garis/hyperplane. Dalam penelitian ini data diuji menggunakan 10 fold cross validation, dimana data dibagi menjadi dua yaitu 9/10 bagian adalah data training dan sisanya 1/10 bagian adalah data testing. Pengujian dilakukan sampai 10 kali dengan kombinasi yang berurutan. Berikut hasil dari klasifikasi dari aplikasi weka menggunakan kernel linear..

```
Time taken to build model: 1.57 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      1079      89.1736 %
Incorrectly Classified Instances    131      10.8264 %
Kappa statistic                    0.6892
Mean absolute error                 0.1083
Root mean squared error             0.3239
Relative absolute error             23.5975 %
Root relative squared error         77.3539 %
Total Number of Instances          1210

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
Weighted Avg.  0,846  0,282  0,915  0,846  0,890  0,691  0,832  0,907  Pos
                0,718  0,054  0,805  0,718  0,759  0,691  0,832  0,645  Neg

```

Gambar 11 Hasil klasifikasi kernel linear

Dari gambar 11 menunjukkan hasil sistem kernel linear dalam pengelompokan klasifikasi positif dan negatif yang berhasil teridentifikasi. Jumlah kata positif yang benar adalah 873, sedangkan negatif berjumlah 206. Pada penelitian sebagai pembanding hasil klasifikasi menggunakan kernel non linear polynomial sebagai bahan evaluasi mana kernel yang memiliki akurasi yang baik. Kernel non linear dipakai didalam situasi data hanya dapat dipisahkan dengan garis lengkung pada ruang dimensi tinggi. Hasil klasifikasi menggunakan kernel polynomial .

```
Time taken to build model: 2.79 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      1021      84.3802 %
Incorrectly Classified Instances    189      15.6198 %
Kappa statistic                    0.4743
Mean absolute error                 0.1542
Root mean squared error             0.3952
Relative absolute error             43.1346 %
Root relative squared error         92.9132 %
Total Number of Instances          1210

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
Weighted Avg.  0,844  0,454  0,845  0,844  0,822  0,521  0,695  0,758  Pos
                0,411  0,022  0,855  0,411  0,555  0,521  0,695  0,491  Neg

```

Gambar 12 Hasil kernel Polynomial

Dari gambar 12 kernel polynomial berhasil melakukan klasifikasi positif sebanyak 903, sedangkan untuk label negatif sebanyak 118. Dari kedua kernel akan dihitung akurasi, presisi dan recall untuk mengetahui mana kernel yang cukup baik dalam klasifikasi data sentimen Go-Pay.

6. Evaluasi Hasil

Evaluasi hasil dilakukan untuk mengukur dan dapat menarik kesimpulan dari proses penelitian yang telah dilakukan. Pengukuran kinerja diperlukan untuk menggambarkan seberapa baik sistem dalam proses klasifikasi data. Dalam penelitian banyak digunakan metode untuk evaluasi hasil salah satunya *Confusion Matrix*. Dari *Confusion Matrix* akan dihitung akurasi, presisi dan recall untuk membandingkan kernel mana yang cukup baik. Berikut Tabel 8 *Confusion Matrix* dari kedua kernel.

Tabel 8 Confusion Matrix

Prediksi	Linear		Polynomial	
	Positif	Negatif	Positif	Negatif
Positif	873	50	903	20
Negatif	81	206	169	118

Dari data *Confusion Matrix* maka kernel linear dapat dihitung akurasi, presisi dan recall sebagai berikut.

$$\text{Akurasi} = \frac{873+206}{873+206+81+50} * 100\% = 89,17\%$$

$$\text{Presisi} = \frac{873}{873+81} * 100\% = 91,5\%$$

$$\text{Recall} = \frac{873}{873+50} * 100\% = 94,58\%$$

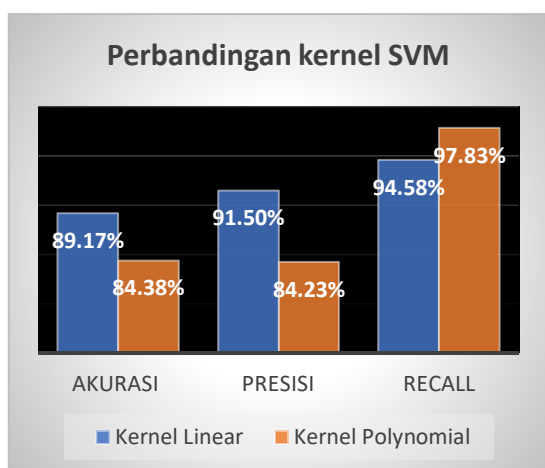
Sedangkan untuk hasil dari kernel polynomial sebagai berikut.

$$\text{Akurasi} = \frac{903+118}{903+118+169+20} * 100\% = 84,38\%$$

$$\text{Presisi} = \frac{903}{903+169} * 100\% = 84,23\%$$

$$\text{Recall} = \frac{903}{903+20} * 100\% = 97,83\%$$

Setelah mendapat hasil perbandingan kernel SVM dapat dilihat di gambar diagram pada Gambar 13 .



Gambar 13 Hasil Perbandingan Kernel SVM

Dari hasil diatas yang didapat akurasi kernel linear mendapatkan hasil akurasi sebesar 89,17%, presisi sebesar 91,5% dan recall 97,83%, sedangkan untuk kernel polynomial mendapatkan akurasi 84,38%, presisi 84,23% dan recall 97,83%. Metode klasifikasi SVM menggunakan kernel linear cukup baik dalam akurasi dan presisi, namun untuk recall kernel polynomial cukup tinggi dibandingkan kernel linear. Walaupun tingkat akurasi dari polynomial lebih rendah daripada linear, perbedaannya tidak terlalu jauh.

## E. PENUTUP

### Kesimpulan

Berdasarkan hasil dari penelitian, maka dapat disimpulkan sebagai berikut.

1. Analisis sentimen tentang Go-Pay untuk mengetahui ulasan komentar pengguna Go-pay di media sosial Twitter didapatkan dari pelabelan sentimen menggunakan *lexicon based* bahwa

pengguna banyak berkomentar positif yang berjumlah 923 ulasan,se sedangkan untuk negatif berjumlah 287.

2. Dari penelitian ini juga menunjukkan metode klasifikasi SVM dengan membandingkan kernel cukup baik. Untuk kernel linear mendapatkan akurasi yang cukup baik sebesar 89,17% dimana sistem kernel dapat melakukan klasifikasi dengan sebanyak 1109 ulasan. Sedangkan untuk kernel polynomial mendapatkan akurasi sebesar 84,38% dan dapat melakukan klasifikasi dengan benar, sebanyak 1021 ulasan. Penelitian ini juga dapat disimpulkan bahwa data twitter untuk ulasan Go-Pay dapat dipisah secara linear cukup baik.
3. Sesuai dengan hipotesa sebelumnya bahwa persentase dari komentar positif dari pengguna gopay lebih tinggi dibandingkan komentar negatifnya yaitu untuk komentar positif menghasilkan 72% dan untuk komentar negatif menghasilkan 17% dari kernel linear SVM.

## Saran

1. Dari pelabelan yang telah didapat, masih ada ulasan yang bersentimen negatif. Untuk perusahaan Gojek yang menaungi Go-Pay bisa sebagai bahan evaluasi untuk meningkatkan kualitas pelayanan agar pengguna mendapatkan kepuasan dari peningkatan pelayanan tersebut.
2. Dari klasifikasi menggunakan *Support Vector Machine* (SVM) mendapatkan akurasi yang cukup baik dengan data yang lebih banyak. Dan untuk penelitian selanjutnya bisa menggunakan metode lainnya sebagai pembanding akurasi seperti Naive Bayes, Neural Network atau lainnya sebagai varisasi penelitian.

## F. DAFTAR PUSTAKA

Josi, A., Abdillah, L. A., & Suryayusra. 2014. Studi, P., Informatika, T., Komputer, F. I., Darma, U. B., ... No, Y. (n.d.). *PENERAPAN TEKNIK WEB SCRAPING PADA MESIN PENCARI ARTIKEL ILMIAH*.

Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal Maret, 1*(1), 32–41. Retrieved from [https://www.researchgate.net/profile/G\\_hulam\\_Buntoro/publication/316617194\\_Analisis\\_Sentimen\\_Calon\\_Gubernur\\_DKI\\_Jakarta\\_2017\\_Di\\_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf](https://www.researchgate.net/profile/G_hulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf)

Yosmita Praptiwi, D. (2018). *Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy*.

Dailysocial 2018 : Fintech Report. (2018). Website : <https://dailysocial.id/report/post/fintech-report-2018>

Tirto 2019 : Evolusi gojek sebagai fintech (2019). Website : <https://tirto.id/evolusi-gojek-sebagai-fintech-lewat-go-pay-cAvw>

APJII 2018 : Penetrasi & Perilaku Pengguna Internet Indonesia (2018). Website : <https://apjii.or.id/survei2018>

Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *E-Proceeding of Engineering, 2*(1), 1–7.

Github : Id Stopwords. Website : <https://github.com/masdevid/ID-OpinionWords>

Github : kamus. Website : <https://github.com/riochr17/Analisis-Sentimen-ID/tree/master/kamus>

Himawan, H., Kaswidjanti, W., Sentimen, A., Sosial, M., & Based, L. (2018). *Metode Lexicon Based Dan Support Vector Machine Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit. 2018*(November), 235–244.

Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 1*(12), 1725–1732. Retrieved from <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>