

# Analisis Sentimen Pengguna Sistem Pay Later Menggunakan Support Vector Machine Metode Pembobotan Lexicon

Ferra Junian Wahidna<sup>1</sup>, Paramitha Nerisafitra<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika Fakultas Teknik Universitas Negeri Surabaya

<sup>1</sup>[ferra.18021@mhs.unesa.ac.id](mailto:ferra.18021@mhs.unesa.ac.id)

<sup>2</sup>[paramithanerisafitra@unesa.ac.id](mailto:paramithanerisafitra@unesa.ac.id)

**Abstrak**— Layanan *pay later* sangat mudah dengan cepat populer di masyarakat, hal ini disebabkan karena fitur ini cenderung mudah digunakan dan populer karena tertanam pada *e-commerce*. Banyaknya penyedia layanan *pay later* menyebabkan diperlukannya pemilahan penyedia *paylater* mana yang akan digunakan berdasarkan berbagai pertimbangan termasuk *review* dari pengguna lain. Di sisi lain, *review* pengguna lain dapat diperoleh dari Twitter. Data dari Twitter menunjukkan terdapat ribuan *tweets* pada tahun 2021 hingga 2022 berisikan opini masyarakat terkait penggunaan *pay later*. *Tweets* tersebut membahas tentang keunggulan, keluhan, dan ulasan dari penggunaan *pay later*. Namun, beberapa keunggulan, keluhan, dan ulasan tersebut banyak yang bersifat abstrak sehingga masih belum optimal pemanfaatannya. Penelitian ini bertujuan untuk mengklasifikasikan data *tweet* yang berkaitan dengan *pay later* menggunakan algoritma *Support Vector Machine* (SVM). Dari penelitian ini telah berhasil dibangun model klasifikasi SVM untuk kasus sentimen Shopee Paylater dan Go Paylater. Pada pemodelan sentimen Shopee Paylater diperoleh bahwa model telah dapat memprediksi kelas data uji dengan akurasi 89.74%. Pada pemodelan sentimen Go Paylater diperoleh bahwa model telah dapat memprediksi kelas data uji dengan akurasi 90.27%.

**Kata Kunci** : Analisis Sentimen, Klasifikasi, *Pay later*, SVM, Lexicon

## I. PENDAHULUAN

*Financial technology* atau disingkat *fintech* adalah inovasi di bidang jasa keuangan yang tengah naik daun di Indonesia beberapa tahun terakhir. Dalam praktiknya, *fintech* tidak terlepas dari kebijakan pemerintah, dibuktikan dengan adanya Pasal 5 Undang-Undang Nomor 21 Tahun 2011 Tentang Otoritas Jasa Keuangan (OJK) yang menyatakan bahwa OJK berfungsi menyelenggarakan sistem pengaturan dan pengawasan yang terintegrasi

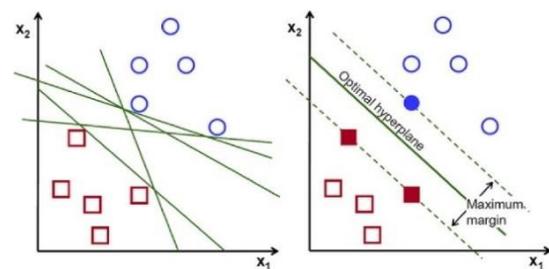
terhadap keseluruhan di dalam sektor jasa keuangan. Salah satu inovasi terkait *fintech* tersebut adalah system *pay later*, yaitu sebuah fitur untuk melakukan metode pembayaran dengan cicilan tanpa kartu kredit saat melakukan pembelian melalui suatu *platform*. Layanan *pay later* sangat mudah dengan cepat populer di masyarakat, hal ini disebabkan karena fitur ini cenderung mudah digunakan dan populer karena tertanam pada *e-commerce*. Dalam memanfaatkan fitur ini, masyarakat dituntut untuk selalu bijaksana karena mudahnya menggunakan fitur ini dapat menyebabkan *overload* pengeluaran pada waktu yang akan datang [1].

Banyaknya penyedia layanan *pay later* di masyarakat menyebabkan munculnya kebingungan masyarakat dalam memilih dan memilih layanan yang baik. Beberapa penyedia jasa *pay later* tidak memiliki izin dari Otoritas Jasa Keuangan atau bisa disebut *illegal* dan terkadang memiliki suku bunga yang tinggi, hal inilah yang menyebabkan munculnya keraguan masyarakat dalam memilih layanan *pay later*. Data dari Twitter menunjukkan terdapat ribuan *tweets* pada tahun 2021 hingga 2022 berisikan opini masyarakat terkait penggunaan *pay later*. Ini membuktikan bahwa *review* masyarakat tentang *pay later* masih menjadi isu sensitif di masyarakat. Beberapa isu pada *tweet* tersebut membahas tentang keunggulan, keluhan, dan ulasan dari penggunaan *pay later*. Namun, beberapa keunggulan, keluhan, dan ulasan tersebut banyak yang bersifat abstrak sehingga masyarakat merasa kebingungan dalam menentukan *tweet* tersebut termasuk ulasan positif atau negatif. Berdasarkan hal tersebut, diperlukan suatu metode untuk mengklasifikasikan setiap *tweet* mengenai *pay later*. Data dari twitter tersebut nantinya akan dianalisa dan diklasifikasi menggunakan suatu metode untuk memberikan bobot tertentu yang nantinya dapat mengelompokan sentiment tersebut.

Dalam mempelajari dan menganalisa pendapat atau *review* konsumen terhadap suatu sistem maka diperlukan suatu instrumen salah satunya adalah *Sentiment analysis*. *Sentiment analysis* atau dalam Bahasa Indonesia analisis sentimen merupakan suatu metode untuk menganalisa dan mengekspresikan suatu *statement*, sikap, pilihan atau sebuah bentuk emosi berdasarkan penilaian tertentu terhadap suatu tulisan [2]. Analisis sentimen dapat diimplementasikan dalam berbagai aspek kehidupan manusia diantaranya adalah bidang medis, pemerintahan, pendidikan maupun bidang bisnis. Tujuan utama dari analisis ini adalah untuk membantu dalam mengambil keputusan maupun suatu kebijakan [3]. Hasil analisa dari *sentiment analysis* adalah berupa *data text* yang telah dikelompokkan pada beberapa kelas, umumnya digunakan kelas berupa katagori positif, netral, negatif ataupun ekspresi lain. Hasil lain yang diberikan analisis ini adalah hasil representasi secara visual terhadap data yang berupa frekuensi kata yang sering digunakan (*Wordcloud*) maupun grafik atau tabel tertentu [4]. Pembobotan yang digunakan pada penelitian ini adalah *Lexicon Based Features*. *Lexicon* merupakan metode pengelompokan sentimen yang bernilai positif atau negatif yang didasarkan pada sebuah database atau kamus (*lexicon*) yang berisi *dictionaries* yang telah diberi nilai bobot tiap katanya, kemudian setiap kata pada sentimen akan memberikan nilai bobot tersendiri dan akan dijumlahkan. *Lexicon* dapat membantu melabeli kumpulan data di *Support Vector Machine*. Penentuan nilai sentimen akan dilakukan pada langkah selanjutnya dengan mengklasifikasikan menjadi beberapa bagian berdasarkan nilai yang ditetapkan. Setelah fitur pembobotan kata dan *Lexicon Based Features* diperoleh, maka data-data tersebut akan diklasifikasi dengan metode *Support Vector Machine* [5]. Kombinasi ini dilakukan secara berurutan dengan penggunaan *Lexicon* untuk menentukan nilai sentimen kemudian data tersebut digunakan sebagai data pelabelan untuk *Support Vector Machine*. Kombinasi ini menjadikan metode *Lexicon* sebagai sarana untuk mentransfer pembelajaran ke *Support Vector Machine*. Tujuan dikombinasikannya metode dan algoritma ini adalah agar proses pelabelan sentimen dapat mendapatkan akurasi yang baik [6].

Terdapat berbagai macam metode klasifikasi *text mining* dalam melakukan analisis sentimen diantaranya adalah *Support Vector Machine*, *Naive Bayes Classifier*, dan *K-Nearest Neighbor*. Pada penelitian Pamungkas [7], yang membandingkan kinerja algoritma klasifikasi untuk Studi Tanggapan Masyarakat Indonesia terhadap Pandemi Covid-19 pada Media Sosial Twitter menyebutkan bahwa nilai akurasi tertinggi didapatkan dari metode *Support*

*Vector Machine* sebesar 90,1%, sedangkan untuk *Naive Bayes Classifier* mendapatkan nilai akurasi sebesar 79,20%, dan *K-Nearest Neighbor* mendapatkan nilai akurasi sebesar 62,10%. Lalu penelitian serupa yaitu *The Influence of Fake Accounts on Sentimen Analysis related to COVID-19 in Indonesia* menyebutkan bahwa algoritma *Support Vector Machine* memiliki nilai akurasi tertinggi sebesar 80,6%, sedangkan algoritma *Naive Bayes Classifier* hanya memiliki nilai akurasi tertinggi sebesar 59% [8]. Berdasarkan kedua penelitian tersebut, penulis dalam penelitian ini memilih menggunakan metode *Support Vector Machine* karena memiliki nilai akurasi yang cukup tinggi dibanding dengan metode yang lain.



Gambar 1. Skema Support Vector Machine

Algoritma *Support Vector Machine* digunakan untuk mencari *hyperplane* terbaik dalam ruang N-dimensi yang secara jelas mengklasifikasikan titik data. *Hyperplane* adalah sebuah fungsi yang digunakan sebagai pemisah antar kelas yang satu dengan yang lain. Fungsi ini digunakan untuk mengklasifikasikan di dalam ruang kelas dimensi yang lebih tinggi. Dalam bentuk 2 dimensi, fungsi yang digunakan untuk mengklasifikasikan antar kelas disebut dengan *line* whereas. Sedangkan fungsi yang digunakan untuk mengklasifikasikan antar kelas dalam bentuk 3 dimensi disebut *plane* similarly [9].

Kelebihan dari algoritma *Support Vector Machine* adalah dapat mengolah data berdimensi tinggi, tanpa mengalami penurunan performa [10]. *Support Vector Machine* dapat digunakan untuk data yang terdistribusi secara teratur atau tidak diketahui distribusinya. *Support Vector Machine* juga dapat mengatasi masalah klasifikasi dan regresi dengan metode linear maupun non linear [7]. Algoritma ini memiliki teknik yang sangat berguna untuk memecahkan masalah kompleks apa pun, yang disebut *kernel*. Selain itu algoritma ini tidak mengalami kondisi *overfitting* dan bekerja dengan baik bila ada indikasi pemisahan yang jelas. Algoritma ini mampu menangani data berdimensi tinggi dan memiliki tingkat akurasi yang lebih baik dan melakukan prediksi lebih cepat [11].

Berdasarkan latar belakang diatas, maka diperlukan penelitian untuk mengklasifikasikan data *tweet* yang

berkaitan dengan *pay later* menggunakan klasifikasi *Support Vector Machine* sehingga pembaca mendapatkan gambaran mengenai opini terkait penggunaan *pay later*

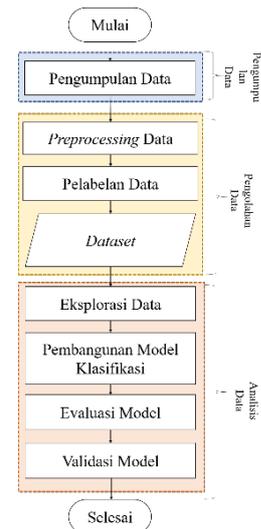
## II. METODE PENELITIAN

### A. Pengumpulan Data dan Sampel

Data diperoleh dengan teknik *crawling* memanfaatkan akses *Application Programming Interface* (API) yang disediakan oleh Twitter. *Crawling* diambil berdasarkan kata “*crawler*” yang memiliki makna memperoleh kumpulan data dimana data yang telah diperoleh tersebut kemudian diolah sesuai dengan kebutuhan pengguna. Secara detail, fungsinya adalah sebagai program aplikasi yang kemudian melakukan memindai halaman internet yang akan diambil datanya [12]. Dengan Twitter API, peneliti memperoleh data Twitter data *tweet* beserta pengguna, waktu, dan kelengkapan lainnya.

Data yang dikumpulkan pada penelitian ini diambil dari media sosial twitter yang merupakan data *tweet* pengguna *pay later*, khususnya *shopee palater* dan *go paylater*. Pada penelitian ini, proses pengambilan data dilakukan dengan mengimplementasikan teknik *crawling* dengan akses API yang diperoleh dari Twitter. Pengimplementasian teknik *crawling* dilakukan dengan bahasa pemrograman R melalui RStudio. Data yang diambil adalah data dengan *keyword* ‘*spaylater*’, ‘*shopee paylater*’, ‘*gopaylater*’, dan ‘*go paylater*’ pada bulan Juli 2022. Dari hasil *crawling* diperoleh data mentah sebanyak 5591 *tweet* dengan *keyword* ‘*spaylater*’, 922 *tweet* dengan *keyword* ‘*shopee paylater*’, 4808 dengan *keyword* ‘*gopaylater*’, dan 401 *tweet* dengan *keyword* ‘*go paylater*’ sehingga diperoleh 6513 *tweet* mentah untuk kelompok Shopee Paylater dan 5209 *tweet* mentah untuk kelompok Go Paylater. Setelah dilakukan *crawling*, data disimpan dalam bentuk *.xlsx*.

Penelitian ini berfokus pada pembangunan model klasifikasi *tweet* yang berkaitan dengan *pay later* dengan mengimplementasikan algoritma klasifikasi *Support Vector Machine* (SVM). Algoritma *Support Vector Machine* (SVM) digunakan pada penelitian ini karena mampu mengidentifikasi *hyperplane* pemisah yang memaksimalkan jarak antara dua kelas yang berbeda [13]. Penelitian dimulai dari identifikasi masalah yang melatar belakangi penelitian, penentuan tujuan yang ingin dicapai pada penelitian, studi literatur pada referensi terkait, pengumpulan data dari Twitter, pengolahan data, penyusunan hasil analisis data, dan penarikan kesimpulan.



Gambar 2. Diagram Alir Metodologi Penelitian

### B. Preprocessing Data

*Preprocessing* data pada *text mining* dilakukan untuk melakukan penyederhanaan pada format data input demi menyederhanakan proses klasifikasi selanjutnya. *Preprocessing* yang dilakukan pada penelitian ini meliputi *cleaning*, *tokenization*, *stopword removal*, *stemming*, dan *slang word replacement*.

### C. Pelabelan Data

Untuk kepentingan pembangunan model, data yang dikumpulkan kemudian dilabeli dengan bantuan kamus leksikon. Kamus leksikon berisi daftar kata positif dan daftar kata negatif. Pada tahap ini, data dilabelkan menjadi tiga kelas yaitu ‘positif’, ‘negatif’, dan ‘netral’ berdasarkan penghitungan atau skoring yang dilakukan. Untuk setiap kata positif yang terkandung pada suatu *tweet* maka skor +1 akan ditambahkan pada *tweet* tersebut, sebaliknya untuk setiap kata negatif yang terkandung pada suatu *tweet* maka skor -1 akan dikurangkan pada *tweet* tersebut. Skor akhir kemudian dihitung dengan menjumlahkan seluruh skor. Jika diperoleh total skor >0 maka *tweet* diklasifikasikan ke dalam kelas positif. Sebaliknya, jika diperoleh total skor <0 maka *tweet* diklasifikasikan ke dalam kelas negative, dan jika total skor = 0 maka diklasifikasikan sebagai netral.

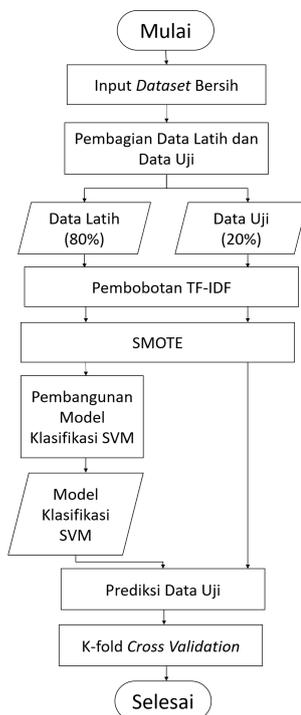
### D. Eksplorasi Data

Sebelum pembangunan model klasifikasi dilakukan, eksplorasi data dilakukan untuk mengetahui pola data yang akan digunakan pada pemodelan. Eksplorasi data yang dilakukan pada penelitian ini dilakukan dengan tujuan utama untuk mendapatkan *insight* perbedaan kata-kata yang menyusun kelas positif dan negatif. Eksplorasi data

dilakukan dengan bantuan alat-alat visualisasi data seperti *bar plot* dan *word cloud*.

#### E. Pembangunan Model Klasifikasi

Pembangunan model klasifikasi dilakukan dengan mengimplementasikan algoritma *Support Vector Machine* (SVM). Pada penelitian ini, terdapat tujuh tahap yang dilakukan dalam pembangunan model klasifikasi. Pertama, penginputan *dataset* yang sudah dibersihkan pada tahap *preprocessing*. Kedua, pembagian *dataset* menjadi data latih dan data uji dengan 3 rasio yaitu 80:20, 70:30, dan 60:40. Ketiga, penghitungan ekstraksi fitur dengan pembobotan TF-IDF. Keempat, mentransformasi data tidak seimbang menjadi seimbang dengan metode SMOTE. Kelima, pembangunan model klasifikasi SVM pada data latih. Keenam, untuk kepentingan evaluasi model, model yang sudah dibangun kemudian digunakan untuk memprediksi data uji. Ketujuh, memvalidasi evaluasi model dengan metode *k-fold cross validation* untuk melihat apakah terjadi *overfitting*. Keseluruhan tahap pembangunan model ini dilakukan dengan bahasa pemrograman *Python* dengan *library* utama yang digunakan adalah 'sklearn'. Adapun pembangunan model dilakukan pada masing-masing *dataset* yaitu Shopee Paylater dan Go Paylater.



Gambar 3. Diagram Alir Pembangunan Model Klasifikasi

#### F. Evaluasi Model

Evaluasi model pada penelitian ini dilakukan untuk mengetahui seberapa baik model yang dibangun dalam memprediksi data uji yang tidak digunakan pada pemodelan. Evaluasi dilakukan dengan cara menghitung nilai akurasi, presisi, *recall*, dan *F1-score* berdasarkan *confusion matrix*.

#### G. Validasi Model

Tahapan ini dibutuhkan untuk mengetahui apakah model sudah benar-benar menghasilkan performa yang terbaik atau tidak, karena akan mudah terjadi bias jika hanya mengandalkan evaluasi model dengan menggunakan metode *holdout* atau membagi data dengan rasio tertentu. Metode untuk melakukan tahapan ini adalah *k-fold cross validation*. Cara kerjanya adalah membagi data menjadi subset data latih dan data uji sebanyak jumlah *k* dan terus melakukan perulangan untuk melatih model sebanyak *k*, sehingga peluang terjadinya *overfitting* menjadi lebih kecil. Penelitian ini menggunakan rentang *k* dari 2 hingga 20, dan dibangun program untuk mencari nilai *k* terbaik.

### III. HASIL DAN PEMBAHASAN

Proses pengambilan data dilakukan dengan mengimplementasikan teknik *crawling* dengan akses API yang diperoleh dari Twitter. Pengimplementasian teknik *crawling* dilakukan dengan bahasa pemrograman R melalui RStudio. Data yang diambil adalah data dengan *keyword* 'spaylater', 'shopee paylater', 'gopaylater', dan 'go paylater' pada bulan Juli 2022. Dari hasil *crawling* diperoleh data mentah sebanyak 5591 *tweet* dengan *keyword* 'spaylater', 922 *tweet* dengan *keyword* 'shopee paylater', 4808 dengan *keyword* 'gopaylater', dan 401 *tweet* dengan *keyword* 'go paylater' sehingga diperoleh 6513 *tweet* mentah untuk kelompok Shopee Paylater dan 5209 *tweet* mentah untuk kelompok Go Paylater.

Variabel yang dikumpulkan di antaranya *user\_id* yang menunjukkan id pengguna twitter, *status\_id* yang menunjukkan id tweet, *created\_at* yang menunjukkan tanggal dan waktu dibuatnya tweet, *screen\_name* yang menunjukkan username akun pembuat tweet, dan *text* yang menunjukkan isi tweet. Tabel I menunjukkan contoh data yang dikumpulkan pada penelitian ini.

TABEL I

DATA YANG DIKUMPULKAN DARI PROSES CRAWLING

user_id	status_id	created_at	screen_name	Text
1451692085900283907	1549636219646214146	20/07/2022 06:18:38	tanderey	Gestun di @dzeliion yu, amanah dan trusted banget 🍌 udah ada testi di pinned 📌  Sung WA 📞 WA: <a href="https://t.co/grO74b4XwO">https://t.co/grO74b4XwO</a> (fastresp)  t. Spaylater Kredivo Akulaku Gopaylater Indodana Gopaycibil BRI Ceria Home Credit  #zonauang <a href="https://t.co/PFURNWozk5">https://t.co/PFURNWozk5</a>
1322345249226674176	1549633258748461056	20/07/2022 05:52:14	chocoreokit	yukk yg mau gestun atau cairin limit spaylater instan langsung cairr cuman scan qr ajaa bisa dm/wa aku 📞 trusted dan amanah 🍌🔥 all nominal fee 8% 🍌🔥 testi bisa cek di pinned hihhi ☐☐ #zonauang #zonaba #zonaba
1322984901096869888	1549059155150852098	18/07/2022 15:50:57	dloey11	@fortinthmoon iyaa ada keterangannya bentuk tagihan gitu deh, tp kalo pake spaylater rugi sih menurutku

A. Preprocessing Data

Dari data yang telah dilakukan *crawling* dapat terlihat bahwa sebagian dari *tweet* yang dikumpulkan adalah *tweet* akun yang menawarkan jasa gestun (gesek tunai) *Paylater* atau berjualan yang tidak mengandung sentimen positif atau negatif sehingga kurang relevan dengan tujuan penelitian. *Tweet* tersebut juga ditemukan sangat banyak karena bersifat iklan sehingga umumnya dilakukan berulang kali. Oleh karena itu, sebelum memasuki tahapan utama *preprocessing*, dilakukan penghapusan *tweet* kurang relevan tersebut sehingga diperoleh 505 *tweet* untuk *dataset* Shopee Paylater dan 180 *tweet* untuk *dataset* Go Paylater.

Tahapan *preprocessing* utama yang dilakukan pada meliputi *cleaning*, *tokenization*, *stopword removal*, *stemming*, dan *slang word replacement*. *Cleaning* dilakukan untuk mengambil teks yang diperlukan untuk pemodelan (menghapus *hashtag*, URL, dan lain-lain) dan mengimplementasikan *case folding*. Gambar 4 menunjukkan pengimplementasian *cleaning*.

no	text
0	1 haaa bingung saya. mending kamu pake spaylater...
1	2 sekarang gw udh gak mau deh beli beli barang y...
2	3 gw juga bener bener harus atur duit banget men...
3	4 @tanyakanrl bertahan hidup karena spaylater
4	5 @dorhwayaki smpe skrg gue gaberani anjir aktif...
...	...
508	509 "shopee paylater pasti baju2 mu?"nneh??? haha...
509	510 social media yang paling nyaman buat gue "twit...

no	text
0	1 haaa bingung saya mending kamu pake spaylater...
1	2 sekarang gw udh gak mau deh beli beli barang y...
2	3 gw juga bener bener harus atur duit banget men...
3	4 bertahan hidup karena spaylater
4	5 smpe skrg gue gaberani anjir aktifin spaylater
...	...
508	509 shopee paylater pasti baju mu heh hahahah pu...
509	510 social media yang paling nyaman buat gue twit...
510	511 kek sumpah yaaa gemes bgt pdhl selama hidupku ...
511	512 km terlalu branded buat ak yg shopee paylater ...
512	513 dek kalau mau lahiran di rumah sakit kita haru...

513 rows x 2 columns  
Gambar 4. Implementasi *Case Folding* (Atas) dan Penghapusan Karakter yang Tidak Diperlukan (Bawah)

Proses *cleaning* kemudian dilanjutkan dengan *tokenizing* dan *stopword removal* untuk menghapus kata-kata yang merupakan *stopwords*. Gambar 5 menunjukkan implementasi *stopword removal* yang mencakup *tokenization*

no	text
0	1 haaa bingung mending kamu pake spaylater aja
1	2 sekarang gw udh gak mau deh beli beli barang y...
2	3 gw bener bener atur duit banget menjadi lebih ...
3	4 bertahan hidup spaylater
4	5 smpe skrg gue gaberani anjir aktifin spaylater
...	...
508	509 shopee paylater baju mu heh hahahah puji tuh...
509	510 social media paling nyaman buat gue twitter w...
510	511 kek sumpah yaaa gemes bgt pdhl selama hidupku ...

Gambar 5. Implementasi *Stopword Removal* yang Mencakup *tokenization*

*Preprocessing* kemudian dilanjutkan dengan *stemming* atau normalisasi kata ke dalam bentuk dasarnya. Gambar 6 menunjukkan implementasi *stemming* pada penelitian ini.

no	text
0	1 haaa bingung mending kamu pake spaylater aja
1	2 sekarang gw udh gak mau deh beli beli barang y...
2	3 gw bener bener atur duit banget jadi lebih iri...
3	4 tahan hidup spaylater
4	5 smpe skrg gue gaberani anjir aktifin spaylater
...	...
508	509 shopee paylater baju mu heh hahahah puji tuhan...
509	510 social media paling nyaman buat gue twitter wh...
510	511 kek sumpah yaaa gemes bgt pdhi lama hidup cmn ...
511	512 km terlalu branded buat ak yg shopee paylater
512	513 dek kalau mau lahir rumah sakit harus lucu sek...

513 rows x 2 columns

Gambar 6. Implementasi Stemming untuk Mengembalikan Kata ke Bentuk Dasarnya

Selanjutnya, untuk mengurangi variasi kata karena perbedaan penulisan atau adanya *slang words*, dilakukan penggantian *slang words* yang diilustrasikan pada Gambar 7.

	sLang	formal
0	woww	wow
1	aminn	amin
2	met	selamat
3	netaas	menetas
4	keberpa	keberapa
...	...	...
15010	gabisa	tidak bisa
15011	menghadeh	aduh
15012	trimakasih	terima kasih
15013	good	bagus
15014	muas	puas

15015 rows x 2 columns

Gambar 7. Implementasi Replace Slang Words untuk mengubah kata yang informal (slang) menjadi kata formal

### B. Pelabelan Data dengan Kamus Lexicon

Setelah melalui tahap *preprocessing*, data yang diperoleh kemudian dilabeli menurut sentimen yang terkandung dengan bantuan kamus leksikon. Pada tahap pelabelan dengan bantuan kamus leksikon, teks diklasifikasikan menjadi tiga kelas yaitu positif (skor > 0), negatif (skor < 0), dan netral (skor = 0). Berdasarkan pelabelan data yang dilakukan, maka diperoleh hasil bahwa pada data Shopee Paylater, 55 teks atau sebanyak 10,8% teks diklasifikasikan positif, 196 teks atau sebanyak 38,8% teks diklasifikasikan negatif, dan 255 sisanya (50,4%) diklasifikasikan netral. Pada data Go Paylater, 34 teks atau sebanyak 18,7% teks diklasifikasikan positif, 61 teks atau sebanyak 33,5% teks diklasifikasikan negatif, dan 87 sisanya (47,8%) diklasifikasikan netral.

### C. Analisis Sentimen Sistem Paylater

Shopee Paylater dan Go Paylater merupakan *platform* yang menyediakan jasa *paylater* yang memungkinkan pembeli untuk mendapatkan barang atau jasa yang diinginkan terlebih dahulu sebelum melunasi pembeliannya. Pengguna kedua *platform paylater* ini banyak mengungkapkan ulasan baik berupa pujian maupun keluhan atas pengalamannya dalam menggunakan *platform paylater* ini. Analisis sentimen system *paylater* dilakukan untuk mengetahui pola untuk masing-masing sentimen sebelum pembangunan model berbasis *machine learning* dilakukan. Pada pembahasan ini, analisis sentimen dibagi menurut *platform* penyedia jasanya.

1) *Analisis Sentimen Pengguna Shopee Paylater*: Kata yang paling banyak disebut adalah ‘spaylater’ karena merupakan topik utama yang dibahas. Kata ‘bayar’ juga merupakan kata yang banyak ditemukan karena pengguna Shopee Paylater banyak yang mengulas pembayaran melalui *platform paylater* ini. Selain itu, kata ‘limit’ juga banyak ditemukan karena *platform paylater* sangat bergantung dengan limit yang disediakan *platform* untuk setiap penggunanya.

Pengguna yang mengungkapkan sentimen positif terhadap Shopee Paylater banyak membahas pembayaran di Shopee, ditandai dengan tingginya frekuensi kata ‘pakai’, ‘shopee’, dan ‘bayar’. Selain itu, kata ‘beli’, ‘mau’, dan ‘yuk’ juga menunjukkan kesan positif terhadap pengalaman berbelanja dengan Shopee Paylater. Tidak sedikit pengguna yang mengajak orang lain untuk menggunakan Shopee Paylater dengan kata ‘yuk’.

Pengguna yang mengungkapkan sentimen negatif terhadap Shopee Paylater banyak membahas pembayaran di Shopee, ditandai dengan tingginya frekuensi kata ‘pakai’, ‘shopee’, dan ‘bayar’. Akan tetapi, pada sentimen ini kata ‘tidak’ yang menunjukkan negasi banyak ditemukan. Selain itu, pada sentimen ini, banyak ditemukan kata ‘limit’ yang menunjukkan bahwa pengguna yang mengungkapkan sentimen negatif karena mempermasalahkan limit yang disediakan.

Pengguna yang mengungkapkan sentimen netral terhadap Shopee Paylater memiliki kata atau pembahasan dari dua sentimen, yaitu sentimen negatif dan sentimen positif. Selain itu, pada sentimen ini, banyak ditemukan kata ‘pakai’, ‘shopee’, ‘bayar’ yang menunjukkan bahwa banyak pembahasan terkait pembayaran, dan kata ‘limit’ yang menunjukkan bahwa pengguna yang mengungkapkan sentimen netral juga mempermasalahkan limit yang disediakan.

2) *Analisis Sentimen Pengguna Go Paylater*: Kata yang

paling banyak disebut adalah ‘gopaylater’ karena merupakan topik utama yang dibahas. Kata ‘bayar’ juga merupakan kata yang banyak ditemukan karena pengguna Go Paylater banyak yang mengulas pembayaran melalui *platform paylater* ini. Selain itu, kata ‘promo’, ‘diskon’, ‘cashback’ juga banyak ditemukan karena *platform* ini banyak menawarkan promo, diskon, dan *cashback* untuk menarik pengguna baru.

Berdasarkan analisis sentimen yang dilakukan, dapat disimpulkan bahwa baik pengguna Shopee *Paylater* atau Go *Paylater*, keduanya banyak membicarakan pembayaran yang dilakukan pada masing-masing *platform* di Twitter. Pemberian promo, diskon, dan *cashback* banyak memperoleh tanggapan positif dari para pengguna. Pada kedua *platform* juga terdapat banyak pengguna mengeluhkan soal limit pinjaman yang diberikan. Topik pembahasan pembayaran, limit pinjaman dan promo juga turut dibahas oleh pengguna yang mengungkapkan tanggapan netral pada kedua platform di Twitter.

#### D. Pembangunan Model Klasifikasi

Untuk kepentingan pembangunan model, dilakukan pembagian *dataset* menjadi data latih dan data uji. Pada penelitian ini dialokasikan rasio 80:20, 70:30, dan 60:40. Pada rasio pembagian 80:20, 80% data untuk menjadi data latih yang akan digunakan untuk melatih model dan 20% data untuk menjadi data uji yang akan digunakan untuk menguji performa model. Pembagian dilakukan pada masing-masing *dataset* Shopee *Paylater* dan Go *Paylater*.

Sebelum dilakukan pembangunan model klasifikasi, ekstraksi fitur dilakukan untuk mengubah data yang berupa teks menjadi angka yang dapat dikenali model. Ekstraksi fitur pada penelitian ini dilakukan dengan metode TF-IDF. Tahapan ini merupakan salah satu tahapan yang sangat penting karena model *machine learning* yang disediakan ‘*sklearn*’ hanya dapat mengenali *input* berupa angka. Selain pengubahan data teks menjadi numerik dengan ekstraksi fitur mengimplementasikan metode TF-IDF, *label encoder* juga dilakukan untuk mengubah label yang mulanya berupa teks ‘negatif’, ‘netral’, dan ‘positif’ menjadi numerik 0, 1, dan 2. Data latih yang telah melewati proses pembobotan dan *label encoder* dicek terlebih dahulu apakah termasuk data tidak seimbang atau seimbang.

SMOTE membuat data sintetik pada kelas minoritas dan akhirnya jumlah kelas minoritas menjadi sama dengan jumlah kelas mayoritas. Terdapat library pada *python* yang harus digunakan yaitu “*from Imblearn.over\_sampling import SMOTE*”. Tabel II dan III menunjukkan perbandingan hasil sebelum dan sesudah SMOTE setelah data latih dibagi menjadi 80% pada dataset Shopee *Paylater* dan Go

*Paylater*.

TABEL II

RINGKASAN HASIL SMOTE PADA DATASET SHOPEE PAYLATER

Kelas	Sebelum SMOTE	Sesudah SMOTE
Positif	10,9%	33,3%
Netral	50,8%	33,3%
Negatif	38,3%	33,3%
Total	100%	100%

TABEL III

RINGKASAN HASIL SMOTE PADA DATASET GO PAYLATER

Kelas	Sebelum SMOTE	Sesudah SMOTE
Positif	19%	33,3%
Netral	47%	33,3%
Negatif	34%	33,3%
Total	100%	100%

Data yang sudah siap kemudian digunakan untuk pembangunan model. Pada penelitian ini dibangun model *Support Vector Machine* (SVM) untuk dapat mengklasifikasikan suatu teks masuk ke dalam kelas negative, netral atau positif. Langkah selanjutnya dicari terlebih dahulu pasangan parameter terbaik. Pencarian parameter terbaik diterapkan pada tiap rasio pembagian data dan pada dua dataset. Tabel IV menunjukkan ringkasan hasil *grid search*.

TABEL IV

HASIL PARAMETER TERBAIK PADA MODEL SVM

Parameter	Shopee <i>Paylater</i>			Go <i>Paylater</i>		
	80:20	70:30	60:40	80:20	70:30	60:40
Kernel	RBF	Linear	RBF	Linear	Linear	<i>Polynomial</i>
Cost value (C)	1	1	1	5	1	2
Degree (p)	3	3	3	3	3	2

Pada penelitian ini dibangun model klasifikasi *Support Vector Machine* (SVM) dengan menggunakan dua dataset dan tiga rasio pembagian data, sehingga menghasilkan enam model SVM yang berbeda.

#### E. Evaluasi Performa Model Klasifikasi

Model klasifikasi yang dibangun kemudian dievaluasi untuk mengetahui performanya. Evaluasi dilakukan pada data uji untuk mengetahui seberapa baik model yang digunakan dalam memprediksi data lain yang tidak diikutsertakan dalam pemodelan. Selain itu, evaluasi juga dilakukan pada data latih untuk melihat apakah model yang dibangun sudah terbebas dari indikasi *overfit* atau model hanya dapat dengan baik memprediksi data latih.

1) *Evaluasi Performa Model Klasifikasi Dataset Shopee Paylater*: Pada data uji dilakukan evaluasi performa model klasifikasi *dataset* Shopee *Paylater* sehingga dapat ditunjukkan *confussion matrix*. Diketahui bahwa nilai akurasi, presisi, *recall*, dan *F1-score* terbaik yang diperoleh pada data uji menggunakan rasio 80:20. Selisih yang jauh menunjukkan bahwa model mengalami *overfitting* sehingga untuk mengatasinya dapat menggunakan *k-fold cross validation*.

TABEL V

EVALUASI MODEL DENGAN DATASET SHOPEE PAYLATER

Model Shopee Paylater	80:20		70:30		60:40	
	Data Uji	Data Latih	Data Uji	Data Latih	Data Uji	Data Latih
Akurasi	81,37%	100%	73,68%	100%	74,38%	100%
Presisi	74,62%	100%	74,25%	100%	72,25%	100%
Recall	81,37%	100%	73,68%	100%	74,38%	100%
F1-Score	77,28%	100%	72,31%	100%	70,9%	100%

2) *Evaluasi Performa Model Klasifikasi Dataset Go Paylater*: Sebagian besar data telah diklasifikasikan pada kelas yang benar. Semua data latih telah diklasifikasikan pada kelas yang benar. Nilai akurasi, presisi, *recall*, dan *F1-score* yang diperoleh lebih rendah dibandingkan menggunakan dataset Shopee *Paylater*, dan nilai metrik terbaik diperoleh dengan membagi data sebanyak 70:30. Selain itu, *gap* atau selisih antara nilai yang diperoleh pada evaluasi diantara data uji dan data latih memiliki selisih yang jauh sehingga terdapat indikasi terjadinya *overfitting* pada model.

TABEL VI

EVALUASI MODEL SVM DENGAN DATASET GO PAYLATER

Model Go Paylater	80:20		70:30		60:40	
	Data Uji	Data Latih	Data Uji	Data Latih	Data Uji	Data Latih
Akurasi	69,44%	100%	70,37%	100%	68,05%	100%
Presisi	63,32%	100%	70,88%	100%	66,67%	100%
Recall	69,44%	100%	70,37%	100%	68,05%	100%
F1-Score	66,24%	100%	69,38%	100%	63,93%	100%

Tabel V dan tabel VI menunjukkan indikasi *overfitting*, sehingga validasi model menggunakan *k-fold cross validation* diperlukan pada penelitian ini.

#### F. Validasi Model

Penelitian ini juga mencari nilai *k* terbaik dengan mencoba semua nilai *k* dari rentang 2 hingga 20. Tujuannya adalah untuk mendapatkan nilai metrik yang maksimal pada model.

#### 1) Validasi Performa Model Klasifikasi Dataset Shopee

*Paylater*: Nilai *k* pada setiap rasio pembagian data memiliki perbedaan, pada rasio 80:20 nilai *k* terbaik adalah 12, pada 70:30 nilai *k* terbaik adalah 13 dan nilai *k* sama dengan 17 merupakan nilai *k* terbaik pada 60:40. Secara keseluruhan nilai metrik berada di atas 80% dan selisih antara data uji dan data latih tidak jauh berbeda sehingga dapat disimpulkan bahwa masalah *overfitting* berhasil diatasi. Tabel VII merupakan tabel ringkasan yang berisikan nilai metrik diantaranya adalah metrik akurasi, metrik presisi, metrik recall, dan metrik *f1*.

TABEL VII

VALIDASI MODEL SVM DENGAN DATASET SHOPEE PAYLATER

Model Shopee Paylater	80:20		70:30		60:40	
	Data Uji	Data Latih	Data Uji	Data Latih	Data Uji	Data Latih
Akurasi	89,74%	89,47%	89%	89,49%	88,16%	89,95%
Presisi	92,25%	90,16%	90%	89,90%	89,74%	90,73%
Recall	89,74%	89,47%	89%	89,49%	88,16%	89,54%
F1-Score	89,39%	89,32%	88,95%	89,95%	87,78%	89,30%

Berdasarkan tabel VII, dapat disimpulkan bahwa hasil klasifikasi model SVM menggunakan dataset Shopee *Paylater* terbaik adalah menggunakan rasio pembagian data sebesar 80:20 dengan nilai *k-fold* sebesar 12.

#### 2) Validasi Performa Model Klasifikasi Dataset Go Paylater

Berbeda dengan Shopee *Paylater*, pada

dataset Go *paylater* dengan rasio pembagian data sebesar 80:20 nilai  $k$  terbaik adalah 18, pada 70:30 nilai  $k$  terbaik adalah 7 dan nilai  $k$  sama dengan 17 merupakan nilai  $k$  terbaik pada 60:40. Hasil metrik akurasi, presisi, *recall*, dan *f1 score* memiliki variasi lebih banyak, tetapi rata-rata berada di atas 75% dan selisih antara data uji dan data latih tidak jauh berbeda sehingga dapat disimpulkan bahwa masalah *overfitting* berhasil diatasi. Tabel VIII merupakan tabel ringkasan yang berisikan nilai metrik diantaranya adalah metrik akurasi, metrik presisi, metrik recall, dan metrik *f1*.

TABEL VIII  
VALIDASI MODEL SVM DENGAN DATASET GO PAYLATER

Model Go Paylater	80:20		70:30		60:40	
	Data Uji	Data Latih	Data Uji	Data Latih	Data Uji	Data Latih
Akurasi	90,27%	85,60%	75%	84,28%	89,07%	75,24%
Presisi	92,28%	88,18%	87%	85,50%	90,85%	82,20%
Recall	90,27%	85,60%	75%	84,28%	89,07%	75,24%
F1-Score	89,84%	85,79%	75,40%	84,19%	89,01%	75,51%

Berdasarkan tabel VIII, disimpulkan bahwa hasil klasifikasi model SVM menggunakan dataset Go *Paylater* terbaik adalah menggunakan rasio pembagian data sebesar 80:20 dengan nilai  $k$ -fold sebesar 18.

#### IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa hasil pelabelan yang dilakukan dengan bantuan kamus leksikon menghasilkan data Shopee *Paylater*, 255 teks atau sebanyak 50,4% teks diklasifikasikan netral, 196 teks atau sebanyak 38,8% teks diklasifikasikan negatif, dan sebanyak 55 teks atau sebanyak 10,8% diklasifikasikan positif. Pada data Go *Paylater*, 87 teks atau sebanyak 47,8% diklasifikasikan sebagai netral, 61 teks atau sebanyak 33,5% teks diklasifikasikan negatif, 34 teks atau sebanyak 18,7% teks diklasifikasikan positif. Hal ini menunjukkan bahwa sentimen pengguna Shopee *paylater* atau Go *paylater* cenderung netral berdasarkan klasifikasi oleh leksikon.

Dari analisis data eksploratif diperoleh bahwa: 1) Baik pengguna Shopee *Paylater* atau Go *Paylater*, keduanya banyak membicarakan pembayaran yang dilakukan pada masing-masing *platform* di Twitter, 2) Pemberian promo, diskon, dan *cashback* banyak memperoleh tanggapan positif dari para pengguna, 3) Pada kedua *platform*, pengguna banyak mengeluhkan soal limit pinjaman yang diberikan, dan 4) Topik pembahasan

pembayaran, limit pinjaman dan promo juga turut dibahas oleh pengguna yang mengungkapkan tanggapan netral pada kedua platform di Twitter.

Dari penelitian ini telah berhasil dibangun model klasifikasi dengan mengimplementasikan algoritma *Support Vector Machine* (SVM) untuk kasus sentimen Shopee *Paylater* dan Go *Paylater*. Pada pemodelan sentimen Shopee *Paylater* diperoleh bahwa validasi model terbaik adalah dengan rasio pembagian data sebesar 80% data latih, dan 20% data uji menggunakan nilai parameter  $C = 1$ ,  $degree = 3$  dan kernel RBF untuk menghasilkan akurasi terbaik sebesar 89,74%. Berbeda dengan dataset shopee *paylater*, pemodelan menggunakan dataset Go *paylater* membutuhkan rasio pembagian data sebesar 80% data latih, dan 20% data uji, menggunakan kernel linear dengan nilai parameter  $C = 5$  dan  $degree = 3$  untuk menghasilkan akurasi terbaik sebesar 90,27%.

#### REFERENSI

- [1] Prastiwi, I. E., & Fitria, T. N. (2021). Konsep Paylater Online Shopping dalam Pandangan Ekonomi Islam. *Jurnal Ilmiah Ekonomi Islam*, 7(1), 425-432.
- [2] Liu, B. (2012) Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5, 1-167.
- [3] Fan, Weiguo, Linda Wallace, Stephanie Rich, and Zhongju Zhang. (2006). "Tapping into the Power of Text Mining". *Communications of the ACM*.
- [4] Gunawan, B., Sastypratiwi, H., & Pratama, E. E. (2018). Sistem Analisis Sentimen Pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jepin (Jurnal Edukasi Dan Penelitian Informatika)*, 4(2), 113-118.
- [5] Bhalerao, K. V. (2021). Customer Reviews Sentiment Analysis: A hybrid technique of Lexicon and Machine Learning based Classification model (SVM, NB, Logistic Regression) (Doctoral dissertation, Dublin, National College of Ireland).
- [6] Muhammadiyah, Rindu Hafil., Laksana, Tri Ginanjar., Arifa, Amalia Beladina. (2022). Kombinasi Support Vector Machine dan Algoritma Berbasis Lexicon dalam Analisis Sentimen Twitter. *Jurnal Ilmu Komputer dan Informatika* (Vol 8, pp. 2)
- [7] Pamungkas, F. S., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia terhadap Pandemi Covid-19 pada Media Sosial Twitter. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 4, pp. 628- 634).
- [8] Pratama, R.P., & Tjahyanto, A. (2022). The Influence of Fake Accounts on Sentimen Analysis related to COVID-19 in Indonesia. *Procedia Computer Science*, 197, 143-150.
- [9] Christian, J., Goldenia, A. E., Chairunnisa, C., Harisa, H., & Prasvita, D. S. (2021). Implementasi Algoritma Support Vector Machine dalam Memprediksi Keberhasilan Suatu Startup Berdasarkan Status Akuisisi. *Senamika*, 2(2), 31-40.
- [10] Purnamawan, I. K. (2015). Support Vector Machine pada Information Retrieval. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 12, 173-180.
- [11] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown

- Pemerintah Jakarta Menggunakan Algoritma Svm. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 31-37.
- [12] Saputra, S. A., Sukur, A., Tangkudung, J., Dlis, F., & Widiastuti, W. (2019). The experiment of physical education teaching style on different categories of student confidence to improve the swimming learning value (Experimental method using design of treatment by factorial 3x2) at Higher School Of Teacher Training And Pedagogy Kusumanegara Jakarta. *Journal of Education, Health and Sport*, 9(6), 98-120.
- [13] Chou, J. S., Cheng, M. Y., Wu, Y. W., & Pham, A. D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications*, 41(8), 3955- 3964.