

HASIL CEK_Analisis Sentimen; Hashtag; Twitter; Support Vector Machine

by Rusydi Umar, Sunardi, Muhammad Nur Ardhiansyah Penerapan
Algoritma Support Vector Machine

Submission date: 05-Dec-2022 12:16PM (UTC+0700)

Submission ID: 1971663289

File name: upport_Vector_Machine_Pada_Analisis_Sentimen_Hashtag_Twitter.pdf (239.84K)

Word count: 4251

Character count: 27031

Penerapan Algoritma Support Vector Machine Pada Analisis Sentimen Hashtag Twitter

Rusydi Umar¹, Sunardi², Muhammad Nur Ardhiansyah^{1,*}

¹Program Studi Magister Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

²Program Studi Teknik Elektro, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Email: ¹ rusydi_umar@rocketmail.com, ² sunardi@mti.uad.ac.id, ^{3,*} muhammad1808048033@webmail.uad.ac.id

Email Penulis Korespondensi: muhammad1808048033@webmail.uad.ac.id

Submitted 21-09-2022; Accepted 31-10-2022; Published 31-10-2022

Abstrak

Berkembangnya industri kreatif di Indonesia ditandai dengan munculnya content creator seperti youtuber dan selebgram. Dengan munculnya content creator menuntut orang berprofesi tersebut harus lebih kreatif dan memunculkan hal-hal baru sesuai dengan trend masyarakat sesuai dengan hukum yang berlaku di Indonesia. Twitter merupakan salah satu media sosial yang dapat digunakan untuk melakukan sharing secara online dan memberikan informasi sesuai dengan trend yang berlaku di masyarakat. Tagar atau hashtag pada twitter sering digunakan pengguna untuk menambahkan pada komentar sehingga ketika hashtag banyak digunakan akan menjadi trending topic. Dengan melakukan analisis sentimen pada trending topic maka akan didapatkan kecenderungan positif dan negatif yang dapat membantu profesi content creator untuk membuat konten. Penelitian ini menggunakan metode Support Vector Machine (SVM) untuk melakukan proses analisis sentimen tentang hashtag twitter. Data yang digunakan adalah data hashtag pada twitter. Data yang telah didapatkan kemudian dilakukan text preprocessing, kemudian pembobotan, dan yang terakhir analisis sentimen menggunakan SVM. Penelitian ini menggunakan data tweet yang diambil pada tanggal 29 Oktober 2022 sebanyak 21 data yaitu BeliPulsaDiGrab yang mempunyai dua kelas sentimen, yaitu sentimen positif sebanyak 99% dan negative 1% data.

Kata Kunci: Analisis Sentimen; Hashtag; Twitter; Support Vector Machine

Abstract

The development of the creative industry in Indonesia is marked by the emergence of content creators such as YouTubers and celebrities. With the emergence of content creators, people in these professions must be more creative and come up with new things in accordance with community trends in accordance with applicable laws in Indonesia. Twitter is one of the social media that can be used to share online and provide information in accordance with the prevailing trends in society. Hashtags or hashtags on Twitter are often used by users to add comments so that when the hashtag is used a lot it will become a trending topic. By analyzing sentiment on trending topics, positive and negative tendencies will be obtained that can help the content creator profession to create content. This study uses the Support Vector Machine (SVM) method to conduct a sentiment analysis process about the twitter hashtag. The data used is hashtag data on twitter. The data that has been obtained is then carried out by text preprocessing, then weighting, and finally sentiment analysis using SVM. This study uses tweet data taken on October 29, 2022 as many as 21 data, namely "BeliPulsaDiGrab" which has two sentiment classes, namely positive sentiment as much as 99% and negative 1% data.

Keywords: Sentiment Analysis; Hashtag; Twitter; Support Vector Machine

1. PENDAHULUAN

Bertumbuhnya industri kreatif di Indonesia berimbas pada perekonomian masyarakat yang lebih baik. Kemajuan teknologi tidak hanya memberikan dampak baik tetapi juga dapat memberikan dampak buruk juga yaitu banyaknya kejahatan yang berkaitan dengan aplikasi internet (Rochmadi et al., 2017). Salah satu industri kreatif yaitu menjadi youtuber dan selebgram, keduanya membutuhkan kreativitas dan hal-hal yang baru agar bisa menghasilkan konten menarik, terbaru, dan berkualitas serta digemari oleh viewers. Kreativitas sangat dibutuhkan karena persaingan semakin meningkat dan keharusan mengikuti trend agar tidak tertinggal. Twitter yang terhubung pada internet dapat melakukan sharing (Bintang et al., 2018). Twitter dapat mengikuti trend yang berkembang di masyarakat Indonesia dengan pengguna mencapai 310 juta dan di dunia mencapai 2,31 triliun (Zuhriyanto et al., 2018) dengan kemungkinan akan semakin bertambah disetiap tahunnya. Banyaknya pengguna menjadikan twitter sebagai tempat berkumpulnya data berbasis text termasuk kicauan yang seharusnya tidak pantas untuk dikonsumsi publik. Bertumbuhnya media sosial menyebabkan begitu mudahnya dalam mendapatkan informasi (Rozi et al., 2017) diantaranya adalah informasi mengenai trend yang sedang berkembang di masyarakat.

Analisis Sentimen adalah bidang ilmu yang menganalisa pendapat, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa, topic, dan atribut mereka (Liu, 2010). Analisis sentimen dapat berupa opini-opini untuk mengekspresikan atau mengungkapkan nilai sentimen penambang kontekstual teks mengekstrak informasi subjektif sumber data. Saat ini sudah banyak dilakukan penelitian tentang analisis sentimen diantaranya mengetahui kepuasan mahasiswa dalam proses pendidikan menggunakan angket mahasiswa (Sunardi et al., 2018) dan mengetahui tingkat kepuasan pengguna terhadap penyedia layanan telekomunikasi seluler (Rofiqoh et al., 2017).

Terdapat istilah hashtag atau tagar pada twitter berfungsi untuk mengelompokkan konten dan mempermudah pencarian. Penggunaan hashtag terbanyak akan menjadi trending topic pada twitter. Sosial media bukan hanya untuk mencari teman atau sharing tapi dapat dimanfaatkan untuk melakukan promo dagang atau jual beli sampai dengan promo partai politik (Buntoro, 2017). Data teks yang berkembang di twitter pada umumnya merupakan hasil presentasi suatu hal yang sedang berkembang pada masyarakat (Setyobudi et al., 2018). Dengan begitu trending topic pada twitter merupakan

Copyright © 2022 Rusydi Umar, Page 1607

JURIKOM is licensed under a Creative Commons Attribution 4.0 International License

cerminan dari *trend* yang sedang berkembang di masyarakat. Analisis sentimen *hashtag* twitter dapat digunakan untuk membantu para *content creator* dalam mendapatkan ide untuk menjadi konten, dengan begitu data yang didapat dari twitter yang sedang berkembang dapat diambil dan diolah untuk mendapatkan informasi.

Metode yang digunakan pada penelitian ini menggunakan metode *Support Vector Machine (SVM)* yang merupakan metode untuk melakukan prediksi dan klasifikasi. Konsep klasifikasi SVM adalah mencari *hyperplane* terbaik dan berfungsi sebagai pemisah dua kelas data menggunakan beberapa titik data terpilih. Pada penelitian dilakukan analisis tentang *hashtag* pada twitter.

2. METODOLOGI PENELITIAN

Penelitian terdahulu pernah dilakukan dan terdapat kesamaan objek seperti peneliti (Zuhdi et al., 2019). Memanfaatkan API yang sudah ada pada twitter untuk mendapatkan data dari kedua CAPRES 2019, menghasilkan bahwa masyarakat memiliki opini positif 300, negatif 68, netral 132 untuk paslon 1 dan positif 312, negatif 91, netral 97 untuk paslon 2. Peneliti (Ernawati & Wati, 2018) melakukan penelitian tentang Penerapan Algoritma *K-Nearest Neighbor* pada Analisis Sentimen Review Agen *Travel*. Pengambilan data menggunakan data *review* agen *travel* dari situs https://www.trustpilot.com/categories/travel_holidays, menghasilkan pengklasifikasi k-NN sebanyak *review* 100 positif dan negatif dimana nilai akurasi yang dihasilkan mencapai 87% dengan nilai AUC sebesar 0.916.

(Sunardi et al., 2018) melakukan analisis sentimen berupa data angket mahasiswa untuk mengetahui kepuasan mahasiswa dalam proses pendidikan. Metode yang digunakan adalah NBC dengan diawali membagi dokumen menjadi data latih dan data uji. Hasil klasifikasi metode mencapai tingkat presisi 75%, recall 75%, dan akurasi 80%.

Peneliti (Zuhriyanto et al., 2018) melakukan perbandingan *live forensics* pada keamanan media sosial Instagram, Twitter, dan Facebook. Teknik yang digunakan pada digital forensik yaitu dengan melakukan analisis komputer atau perangkat lunak digital. Analisis digital forensik terbagi menjadi dua kategori yaitu Dead/Tradisional dan Live. Metode yang digunakan yaitu *National Institute of Justice (NIJ)*, kemudian dijalankan menggunakan perangkat lunak (FTK Imager) sebagai bahan pendukung untuk mengetahui keamanan pada aplikasi Twitter.

(Bintang et al., 2018) membuat rancangan untuk perbandingan *live forensics* pada keamanan media sosial Instagram, Facebook dan Twitter di windows 10. Metode yang digunakan yaitu *National Institute of Justice (NIJ)* dengan tahapan berikut *Collection, Examination, Analysis, dan Reporting*. Tujuan penelitian untuk menghasilkan bukti digital forensik yang dapat mengetahui tingkat keamanan pada media sosial Facebook, Instagram dan Twitter.

(Rozi et al., 2012) mengembangkan aplikasi analisis sentiment twitter dengan menggunakan metode Naïve Bayes. Pada penelitian ini menggunakan studi kasus SAMSAT Kota Malang. Hasil dari aplikasi yang dikembangkan dapat melakukan analisis sentiment dimana semakin banyak data latih yang ada maka nilai sentiment yang didapatkan semakin akurat.

(Setianto, 2020) melakukan analisis sentiment twitter dengan tujuan mengetahui dampak yang ditimbulkan pandemi Covid-19 atau virus corona khususnya terhadap tingkah laku masyarakat Jakarta dalam berkomentar di sosial media. Pengujian dilakukan dengan menggunakan analisis TF-IDF Vectorizer untuk mendapatkan feature atau kata-kata berbobot dari tweet/artikel yang akan diujikan analisis sentiment.

Peneliti (Ernawati & Wati, 2018) melakukan Komparasi Teknik Klasifikasi *Teks Mining* pada Analisis Sentimen. Mengabungkan tiga metode *Nive Bayes*, k-NN, dan SVM bertujuan mengetahui performa yang baik dalam hal akurasi. Hasil komparasi SVM lebih unggul dibandingkan dengan Naïve Bayes dan k-NN.

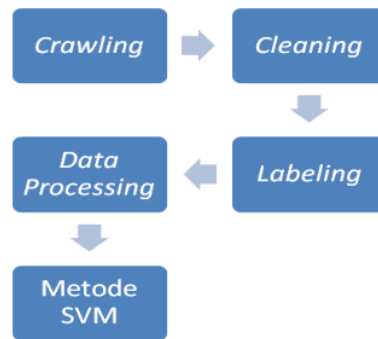
Peneliti (Ipmawati et al., 2017) melakukan Komparasi Teknik Klasifikasi *Teks Mining* pada Analisis Sentimen. Mengabungkan tiga metode *Nive Bayes*, k-NN, dan SVM bertujuan mengetahui performa yang baik dalam hal akurasi. Hasil komparasi SVM lebih unggul dibandingkan dengan Naïve Bayes dan k-NN.

Peneliti (Sari & Wibowo, 2019) melakukan analisis sentiment terhadap pengguna JD.id pada sosial media twitter. Metode yang digunakan adalah NBC dengan pembobotan TF-IDF berbasis konversi ikon emosi (*emoticon*). Hasil pembobotan TF-IDF dan konversi *emoticon* memiliki nilai akurasi sebesar 98%.

Peneliti (Umar et al., 2020) melakukan komparasi metode K-NN dan SVM untuk pencocokan gambar. Tahap dimulai dengan pengolahan data, ekstraksi ciri dan terakhir tahap klasifikasi. Sebanyak 10 gambar yang digunakan terdiri sembilan gambar manipulasi dan satu gambar asli. Pengujian menggunakan aplikasi WEKA dengan menerapkan metode 10-fold *cross validation*. Hasil komparasi bahwa metode K-NN bekerja lebih baik daripada SVM untuk pencocokan gambar.

Objek penelitian yang akan dilakukan adalah penggunaan *hashtag* yang menjadi *trending topic* pada twitter.

Pada penelitian ini menggunakan beberapa tahapan penelitian yaitu *crawling, cleaning, labeling, data processing*, dan implementasi metode SVM. Untuk lebih jelasnya mengenai tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

- Crawling* atau pengambilan data menggunakan API yang sudah disediakan oleh Twitter. Pada penelitian ini mengambil data *hashtag* yang sedang *trending topic* pada twitter kemudian akan melalui proses ekspor kedalam bentuk file *excel*, sehingga didapatkan file *excel* dengan nama “trending”. Didalam file tersebut terdapat beberapa kolom terdiri “timestamp”, “hashtag” dan “tweet”. Setelah itu data dibagi menjadi dua berupa data *training* dan data *testing*..
- Cleaning* adalah proses menghapus data baris yang kosong. Proses ini dilakukan agar tidak ada baris kosong yang diproses sehingga menghemat waktu yang diproses oleh sistem.
- Labeling* merupakan pemberian label pada data yang sudah dilakukan proses *cleaning*. Pada tahap ini data diberi label positif dan negatif dengan metode pengambilan keputusan mempertimbangkan konotasi kata positif untuk label positif dan konotasi negatif untuk label negatif.
- Data Processing* merupakan tahap untuk mendapatkan data yang diinginkan. Tahap *data processing* sangat penting karena untuk mengetahui nilai sentimen dari tweet pengguna twitter, tahapan yang dilakukan dimulai dengan *tokenizing* untuk menguraikan kalimat menjadi kata dan menghilangkan *delir*. Selanjutnya dilakukan proses *stopword removal* dengan cara menghilangkan kosakata yang bukan kata unik dengan menghilangkan kata depan, belakang, kata sambung dan sebagainya. Setelah itu dilakukan *stemming* untuk menguraikan bentuk kata berimbuhan menjadi kata dasar dan kata tidak baku menjadi kata baku.
- Metode SVM yaitu dengan melakukan analisis sentimen menggunakan metode *Support Vector Machine* untuk mengetahui apakah data termasuk sentimen positif atau negatif.

3. HASIL DAN PEMBAHASAN

Penelitian ini mengambil data dengan cara *crawling* memanfaatkan API twitter, mengambil data *tweet* berdasarkan *hashtag* yang sedang *trending topic*. *Crawling* data dapat dilakukan secara terus menerus oleh pengguna pada saat menggunakan aplikasi berbasis web yang terhubung pada internet dengan cara *update*. Hasil pengambilan data dapat dilihat pada Tabel 1 yang merupakan data *hashtag* yang sedang *trending* pada tanggal 29 Oktober 2022 yaitu “#BeliPulsaDiGrab”.

Tabel 1. Hasil *Crawling*

No	Tanggal	Hashtag	Tweet
1	29/10/2022	BeliPulsaDiGrab	#BeliPulsaDiGrab Sangat sangat terbantu banget sama @GrabID Terimakasih buat semua orang yang berada di balik suksesnya apk grab pic.twitter.com/yZhpcKozgU
2	29/10/2022	BeliPulsaDiGrab	Aku td sore habis pake GrabExpress Sameday-Bike untuk ngesend hadiah ke temenku karna lg psbb enggak bisa ketemu2. Ngebantuuu bgt #BeliPulsaDiGrab pic.twitter.com/hkdH2zKaHG
3	29/10/2022	BeliPulsaDiGrab	Terakhir pake grabbike, ada diskon lagi hahaha #BeliPulsaDiGrab pic.twitter.com/RrUNyiX6H5
4	29/10/2022	BeliPulsaDiGrab	So pasti, yg pasti" aku ankny jajan makanan n sering bm lalu order makanan d grabfood krn sering promo n tmpt jajan ny enak" krn mnjlni kehidupan butuh tenaga apa lg yg overthingking mending makan lah biar kuat #BeliPulsaDiGrab pic.twitter.com/QxzXpp7wSK
5	29/10/2022	BeliPulsaDiGrab	Grabcar. Kemarin pagi minta diantar pagi2 liputan video diluar Kota yang bawa kru dan alat byk beudh. Sangat membantu! Thanks #BeliPulsaDiGrab

6	29/10/2022	BeliPulsaDiGrab	Walau dekat karna sy mager jd pesen grab aj wkwk,pengemudinya juga pake masker and nyaranin sy pake helm sendiri,sipp lha #BeliPulsaDiGrab pic.twitter.com/nhPoZii8Ok
7	29/10/2022	BeliPulsaDiGrab	“Grabfood beli makan siang buat calon mertua #BeliPulsaDiGrab”
8	29/10/2022	BeliPulsaDiGrab	Layanan Grab yang terakhir dipakai adalah Grab Bike. Walaupun PSBB tapi tetap masuk kantor, jadi pastinya naik Grab Bike itu praktis banget. Tetap memperhatikan protokol kesehatan, pakai masker dan bawa helm pribadi #BeliPulsaDiGrab
9	29/10/2022	BeliPulsaDiGrab	Grabike kin aja gaesss, anjaayyy #BeliPulsaDiGrab pic.twitter.com/pok28TpzNK
10	29/10/2022	BeliPulsaDiGrab	GrabFood aja dah untuk menu favorite andalah Nasi Telor Pontianak #BeliPulsaDiGrab pic.twitter.com/EI3rJxU6Ys
...
21	29/10/2022	BeliPulsaDiGrab	Nih min.gue pake dalam sehari udah dua kali,btw yang pertama dapat promo dipotong 25% makasih lhooo #BeliPulsaDiGrab pic.twitter.com/wPLyHbG3QA

3.1 Crawling

Dari data yang sudah diperoleh pada Tabel 1 kemudian dilakukan seleksi data dengan memilih kolom yang akan digunakan untuk pelatihan. Kolom yang digunakan adalah kolom “Tweet”.

3.2 Cleaning

Dari kolom yang sudah dipilih kemudian dilakukan proses *cleaning* dengan menghapus baris yang kosong. Hasil dari proses *cleaning* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Cleaning*

Tweet
belipulsadigrab sangat sangat bantu banget sama grabid terimakasih buat semua orang ada balik sukses apk grab pictwittercomyzhpckozgu
aku td sore habis pake grabexpress samedaybike ngesend hadiah temenkuu karna lg psbb enggak ketemu2 ngebantuuu bgt belipulsadigrab pictwittercomhkh2zkahg
akhir pake grabbike diskon hahaha belipulsadigrab pictwittercomrrunyx6h5
so yg aku ankny jajan makan n sering bm lalu order makan d grabfood krn sering promo n tmpt jajan ny enak krn mnjlni hidup butuh tenaga apa lg yg overthingking mending makan lah biar kuat belipulsadigrab pictwittercomqxxpp7wsk
grabcar kemarin pagi minta antar pagi2 liput video luar kota bawa kru alat byk beudh sangat bantu thanks belipulsadigrab
deket karna sy mager jd sen grab aj wkwkpengemudinya pake masker and nyaranin sy pake helm sendiri,sipp lha belipulsadigrab pictwittercomnhpozii8ok
grabfood beli makan siang buat calon mertua belipulsadigrab
layan grab akhir pakai grab bike walaupun psbb tetap masuk kantor jadi naik grab bike praktis banget tetap perhati protokol sehat pakai masker bawa helm pribadi belipulsadigrab
grabike kin aja gaesss anjaayyy belipulsadigrab pictwittercompok28tpznk
grabfood aja dah menu favorite nasi telor pontianak belipulsadigrab pictwittercomei3rjxu6ys
.....
nih mingue pake hari udah kalibtw pertama promo potong 25 makasih lhooo belipulsadigrab pictwittercomwplyhbg3qa
fb ya udah jgn salahin kalau org unfollow bambaang

3.3 Labeling

Setelah didapatkan data yang bersih kemudian dari data awal 100 diberi label positif (1) dan negatif (0) untuk dijadikan *dataset*. Hasil *dataset* yang sudah diberi label kemudian disimpan dalam sistem untuk dijadikan acuan bagi sistem untuk proses analisis sentimen. *Dataset* dapat dilihat pada Tabel 3.

Tabel 3. *Dataset*

Tweet	Label
belipulsadigrab sangat sangat bantu banget sama grabid terimakasih buat semua orang ada balik sukses apk grab pictwittercomyzhpckozgu	1
aku td sore habis pake grabexpress samedaybike ngesend hadiah temenkuu karna lg psbb enggak ketemu2 ngebantuuu bgt belipulsadigrab pictwittercomhkh2zkahg	1
akhir pake grabbike diskon hahaha belipulsadigrab pictwittercomrrunyx6h5	1

grabcar kemarin pagi minta antar pagi2 liput video luar kota bawa kru alat byk beudh sangat bantu thanks 1
belipulsadigrab
deket karna sy mager jd sen grab aj wkwkpengemudinya pake masker and nyaranin sy pake helm sendirisipp 1
lha belipulsadigrab pictwittercomnhpozii8ok

3.4 Data Processing

Pada tahap ini dilakukan *preprocessing* supaya data teks berupa kalimat dapat diproses oleh sistem. Langkah yang dilakukan pada tahap ini adalah:

a. Tokenizing

Tokenizing adalah pemecahan kalimat menjadi beberapa kata berdasarkan kata yang menyusunnya, seperti yang ditunjukkan pada Tabel 4.

Tabel 4. Sampel Proses *Tokenizing*

Tweet
"belipulsadigrab", "sangat", "sangat", "bantu", "banget", "sama", "grabid", "terimakasih", "buat", "semua", "orang", "ada", "balik", "sukses", "apk" "grab" "pictwittercomyzhpcpkozgu" "aku", "td", "sore", "habis", "pake", "grabexpress", "samedaybike" "ngesend", "hadiah", "temenkuu", "karna", "lg", "psbb", "enggak", "ketemu2", "ngebantuuu", "bgt", "belipulsadigrab", "pictwittercomhkdh2zkahg" "akhir", "pake", "grabbike", "diskon", "hahaha", "belipulsadigrab", "pictwittercomrunyix6h5" "grabcar", "kemarin", "pagi", "minta", "antar", "pagi2", "liput", "video", "luar", "kota", "bawa", "kru", "alat", "byk", "beudh", "sangat", "bantu", "thanks", "belipulsadigrab" "deket", "karna", "sy", "mager", "jd", "sen" "grab", "aj", "wkwkpengemudinya", "pake", "masker", "and", "nyaranin", "sy", "pake", "helm", "sendirisipp", "lha", "belipulsadigrab", "pictwittercomnhpozii8ok"

b. Stopword Removal

Tahap selanjutnya adalah melakukan proses *stopword removal* dengan cara menghapus kata umum dan kurang penting seperti kata sambung 'dan', 'atau', 'toh' dan sebagainya, lebih jelasnya dapat dilihat pada Tabel 5.

Tabel 5. Proses *Stopword Removal*

Kalimat Awal	Stopword Removal
"belipulsadigrab", "sangat", "sangat", "bantu", "banget", "sama", "grabid", "terimakasih", "buat", "semua", "orang", "ada", "balik", "sukses", "apk" "grab" "pictwittercomyzhpcpkozgu" "aku", "td", "sore", "habis", "pake", "grabexpress", "samedaybike" "ngesend", "hadiah", "temenkuu", "karna", "lg", "psbb", "enggak", "ketemu2", "ngebantuuu", "bgt", "belipulsadigrab", "pictwittercomhkdh2zkahg" "akhir", "pake", "grabbike", "diskon", "hahaha", "belipulsadigrab", "pictwittercomrunyix6h5" "grabcar", "kemarin", "pagi", "minta", "antar", "pagi2", "liput", "video", "luar", "kota", "bawa", "kru", "alat", "byk", "beudh", "sangat", "bantu", "thanks", "belipulsadigrab" "deket", "karna", "sy", "mager", "jd", "sen" "grab", "aj", "wkwkpengemudinya", "pake", "masker", "and", "nyaranin", "sy", "pake", "helm", "sendirisipp", "lha", "belipulsadigrab", "pictwittercomnhpozii8ok"	"belipulsadigrab", "sangat", "sangat", "bantu", "banget", "sama", "grabid", "terimakasih", "buat", "semua", "orang", "ada", "balik", "sukses", "apk" "grab" "aku", "td", "sore", "habis", "pake", "grabexpress", "samedaybike" "ngesend", "hadiah", "temenkuu", "karna", "lg", "psbb", "enggak", "ketemu2", "ngebantuuu", "bgt", "pake", "grabbike", "diskon", "hahaha", "grabcar", "kemarin", "pagi", "minta", "antar", "pagi2", "liput", "video", "luar", "kota", "bawa", "kru", "alat", "byk", "beudh", "sangat", "bantu", "thanks", "deket", "karna", "sy", "mager", "jd", "sen" "grab", "aj", "wkwkpengemudinya", "pake", "masker", "nyaranin", "sy", "pake", "helm", "sendirisipp", "lha"

c. Stemming

Langkah selanjutnya melakukan proses *stemming* yaitu mengubah semua kata menjadi kata dasar dan menghilangkan imbuhan seperti dapat dilihat pada Tabel 6.

Tabel 6. Proses *Stemming*

Kalimat Awal	Setelah Proses <i>Stemming</i>
"belipulsadigrab", "sangat", "sangat", "bantu", "banget", "sama", "grabid", "terimakasih", "buat", "semua", "orang", "ada", "balik", "sukses", "apk" "grab" "aku", "td", "sore", "habis", "pake", "grabexpress", "samedaybike" "ngesend", "hadiah", "temenkuu", "karna", "lg", "psbb", "enggak", "ketemu2", "ngebantuuu", "bgt",	"beli", "pulsa", "di", "grab", "sangat", "bantu", "banget", "sama", "grab", "terimakasih", "buat", "semua", "orang", "ada", "balik", "sukses", "grab" "aku", "td", "sore", "habis", "pakai", "grab", "express", "kirim", "hadiah", "temenku", "karna",

<p>"pake", "grabbike", "diskon", "hahaha", "grabcar", "kemarin", "pagi", "minta", "antar", "pagi2", "liput", "video", "luar", "kota", "bawa", "kru", "alat", "byk", "beudh", "sangat", "bantu", "thanks",</p> <p>"deket", "karna", "sy", "mager", "jd", "sen", "grab", "aj", "wkwpengemudi", "pake", "masker", "nyaranin", "sy", "pake", "helm", "sendirisipp", "lha"</p>	<p>"lagi", "psbb", "enggak", "ketemu", "ngebantu", "sekali", "pake", "grab", "bike", "diskon" "grab", "car", "kemarin", "pagi", "minta", "antar", "pagi", "liput", "video", "luar", "kota", "bawa", "kru", "alat", "banyak", "sekali", "sangat", "bantu", "thanks", "deket", "karena", "saya", "mager", "jadi", "sen" "grab", "aja", "pengemudi", "pake", "masker", "nyaranin", "saya", "pakai", "helm", "sendiri"</p>
--	---

3.5 Metode SVM

Langkah selanjutnya menghitung nilai similaritas antar kata atau dokumen menggunakan algoritma *Support Vector Machine (SVM)*. Pada penelitian ini menggunakan dua kelas sentimen.

$$\sum \text{lat}K(x_i, x) = \text{constant} \quad (1)$$

Pada Persamaan (1), jika $k(x,y)$ lebih kecil dan nilai y bertambah dan lebih dari nilai x , pada setiap elemen dalam pengukuran tingkat kedekatan titik uji x ke titik x_i dalam *database* yang sesuai. Penelitian ini menggunakan Persamaan (2).

$$D = \{(x_i, y_i) | X_i \in \{-1, 1\}\} i^n \quad (2)$$

y_i adalah 1 atau -1 menunjukkan kelas dari titik x_i , x_i menyatakan vektor nyata p -dimensi. Persamaan (2) merupakan rumus untuk mencari *hyperplane* dari maksimum *margin* yang membagi nilai yang mempunyai nilai $y_i=1$ dari yang mempunyai $y_i=-1$.

$$w \cdot x - b = 0 \quad (3)$$

Persamaan 3 menunjukan dot *product* dan *vector* w merupakan *vector* normal yang tegak lurus dengan *hyperplane*, parameter $b/(||w||)$ menentukan offset *hyperplane* asal sepanjang *vector* normal w . setelah itu untuk pemilihan e dan b untuk memaksimalkan jarak antara *hyperplane* atau memisahkan *margin* tetapi memisahkan data, dijabarkan pada persamaan 4.

$$w \cdot x - b = 1 \text{ dan } w \cdot x - b = -1 \quad (4)$$

Setelah dilakukan pemilihan untuk dua *hyperplane* data yang dilatihkan terpisah secara linier dengan a memisahkan nilai antara *hyperplane* dengan *margin* selanjutnya memaksimalkan jaraknya dengan persamaan $2/(||w||)$. Untuk menghindari titik data (i) jatuh kedalam *margin* maka perlu dirumuskan dengan persamaan 5.

$$w \cdot x_i - b \geq 1 \text{ atau } w \cdot x_i - b \leq -1 \quad (5)$$

$$y_i(w \cdot x_i - b) \geq 1, \text{ untuk semua } 1 \leq i \leq n \quad (6)$$

Untuk mendapatkan optimasi maka minimalkan nilai w dan b ($||w||$) untuk setiap $i=1, \dots, n$ yang dirumuskan pada persamaan 7.

$$y_i(w \cdot x_i - b) - 1 \geq 0 \quad (7)$$

Apabila kelompok atau keluarga *hyperplane* ditemukan, maka dilakukan rumus dengan persamaan 8.

$$y_i(w \cdot x_i - b) - 1 \geq 0 \quad (8)$$

Untuk itu diperlukan semua nilai minimal dengan cara mengirinkan semua a_i untuk semua $+\infty$ sehingga nilai minimum ini akan dicapai oleh semua anggota kelompok. Nilai *constrain* dinyatakan dengan persamaan 9 dengan tujuan mencari pelana atau titik pegangan.

$$\max_{w,b} \max_a \left\{ \frac{1}{2} ||w||^2 - \sum_{i=1}^n a [y_i(w \cdot x_i - b) - 1] \right\} \quad (9)$$

Dengan persamaan 9, semua nilai dapat disajikan dipisahkan sebagai persamaan 10.

$$y_i(w \cdot x_i - b) - 1 > 0 \quad (10)$$

Data yang lebih besar dapat menggunakan matriks kernel untuk menyimpan data pada memori, dengan data yang lebih menggunakan metode *working set* merupakan kumpulan variabel dengan optimasi melalui *current iteration*. Fungsi *kernel* dijabarkan dengan persamaan 11.

$$k(x_i, x_j) = x_i, x_j \quad (11)$$

4. KESIMPULAN

Dari penelitian yang ada dapat dilakukan analisis sentimen twitter menggunakan algoritma SVM dengan pengambilan data memanfaatkan API Twitter. Diperoleh hasil crawling pada tanggal 29 Oktober 2022 terhadap *hashtag* yang sedang *trending* yaitu BeliPulsaDiGrab. Hasil analisis sentimen menunjukkan sebanyak positif 99% negatif 1%.

REFERENCES

- [1] Bintang, R. A. K. N., Umar, R., & Yudhana, U. (2018). Perancangan perbandingan live forensics pada keamanan media sosial Instagram, Facebook dan Twitter di Windows 10. Prosiding SNST Ke-9 Tahun 2018 Fakultas Teknik Universitas Wahid Hasyim, 125-128.
- [2] Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. Integer Journal Maret, 1(1), 32-41. https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907ee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf
- [3] Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. VI(1).
- [4] Ipmawati, J., Kusriani, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. Indonesian Journal on Networking and Security.
- [5] Liu, B. (2010). Sentiment analysis and subjectivity. In Handbook of Natural Language Processing, Second Edition.
- [6] Rochmadi, T., Riadi, I., & Prayudi, Y. (2017). Live Forensics for Anti-Forensics Analysis on Private Portable Web Browser. International Journal of Computer Applications, 164(8). <https://doi.org/10.5120/ijca2017913717>
- [7] 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 Lexicon Based Feature. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya.
- [8] Rozi, I. F., Hadi, S., & Achmad, E. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. 6(1), 37-43.
- [9] Rozi, I. F., Hamdana, E. N., Balya, M., & Alfahmi, I. (2017). PENGEMBANGAN APLIKASI ANALISIS SENTIMEN TWITTER (Studi Kasus SAMSAT Kota Malang). 149-154.
- [10] Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi. 10, 681-686.
- [11] Setianto, W. (2020). Analisis Sentimen Twitter Untuk Mengetahui Dampak Dari Pandemi Covid-19 Terhadap Masyarakat Jakarta. Github, 1-17. <https://github.com/Hyuto/Analisis-Sentimen-Corona-DKI-Jakarta> ABSTRAK
- [12] Setyobudi, W., Alwi, A., & Astuti, I. P. (2018). SENTIMEN ANALISIS TWITTER TERHADAP PENYELENGGARAAN GOJEK TRAVELOKA LIGA 1 INDONESIA. KOMPUTEK. <https://doi.org/10.24269/jkt.v2i1.68>
- [13] Sunardi, Fadlil, A., & Suprianto. (2018). Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Pada Angket Mahasiswa. Saintekbu, 10(2), 1-9. <https://doi.org/10.32764/saintekbu.v10i2.190>
- [14] Umar, R., Riadi, I., & Farook, D. A. (2020). Komparasi Image Matching Menggunakan Metode K-Nearest Neighbor (KNN) dan Metode Support Vector Machine (SVM). Jurnal of Applied Informatics and Computing (JAIC), 4(2), 124-131.
- [15] Zuhdi, A. M., Utami, E., & Raharjo, S. (2019). Analisis Ssentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN. SSRN Electronic Journal, 5, 1-7. <https://doi.org/10.2139/ssrn.3368718>
- [16] Zuhriyanto, I., Yudhana, A., & Riadi, I. (2018). Perancangan Digital Forensik Pada Aplikasi. 2018(November), 86-91.

HASIL CEK_Analisis Sentimen; Hashtag; Twitter; Support Vector Machine

ORIGINALITY REPORT

15%

SIMILARITY INDEX

17%

INTERNET SOURCES

2%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Telkom University Student Paper	8%
2	kpdi12.perpusnas.go.id Internet Source	4%
3	publikasiilmiah.unwahas.ac.id Internet Source	2%
4	repository.uir.ac.id Internet Source	2%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On