

Dendy Aprilianto Nugroho - Analisis Sentimen Terhadap Kegiatan Pembersihan Sampah Pada Media Sosial X Menggunakan Metode Support Vector Machine dan Naïve Bayes

by Layanan Perpustakaan UHAMKA

Submission date: 13-Mar-2024 09:12AM (UTC+0700)

Submission ID: 2319071550

File name: Dendy_Aprilianto_N_TI_TURNITIN_KE-1_-_Dendy_Aprilianto.docx (459.4K)

Word count: 4280

Character count: 28490



Analisis Sentimen Terhadap Kegiatan Pembersihan Sampah Pada Media Sosial X Menggunakan Metode Support Vector Machine dan Naïve Bayes

Dendy Aprilianto Nugroho, Firman Noor Hasan*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, DKI Jakarta, Indonesia

Email: ¹dendyaprilianto98@gmail.com, ²firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Abstrak—Aktivitas kegiatan sehari-hari manusia tidak luput dari sampah, hampir semua kegiatan yang dilakukan manusia akan menimbulkan sampah. Kebiasaan buruk manusia dengan membuang sampah sembarangan yang mengakibatkan pengaruh terhadap keseimbangan lingkungan. Akibat permasalahan ini, ada komunitas anak muda bernama Pandawara Grup yang ingin membantu membersihkan sampah di Pantai Sukabumi, namun aksi tersebut ditolak oleh Kepala Desa dan Karang Taruna setempat yang membuat banyak opini-opini yang menghasilkan sentimen masyarakat yang terdapat pada *tweet* di media sosial X. Karena hal ini, penelitian ini dilakukan agar dapat menganalisis sentimen terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi dengan menggunakan metode *Support Vector Machine* dan *Naïve Bayes*. Tujuan dari dibuatnya penelitian ini yaitu untuk mengetahui sentimen negatif maupun positif masyarakat terhadap permasalahan ini, serta menganalisis perbandingan akurasi dari metode *Support Vector Machine* dan *Naïve Bayes*. Pada analisis sentimen ini dengan menggunakan 2.339 dataset hingga menghasilkan *accuracy* tertinggi pada metode *Support Vector Machine* sebesar 91,67%, sedangkan pada metode *Naïve Bayes* hanya sebesar 63,89%, sehingga ditarik kesimpulan bahwa metode *Support Vector Machine* lebih baik untuk mengklasifikasikan sentimen terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi dibandingkan metode *Naïve Bayes*.

Kata Kunci: Analisis Sentimen; Kegiatan Pembersihan Sampah; Media Sosial X; *Support Vector Machine*; *Naïve Bayes*

Abstract—Human daily activities inevitably produce waste, which negatively impacts environmental balance due to the bad habit of indiscriminately disposing of waste. As a result of this issue, there is a youth community named Pandawara Group that wants to help clean up trash on Sukabumi Beach. However, their initiative faced rejection from the local village chief and youth organization, sparking various opinions on social media platform X. Consequently, this research seeks to analyze public sentiment towards Pandawara Group's waste cleanup efforts at Sukabumi Beach using *Support Vector Machine* and *Naïve Bayes* methods. The objective is to gauge positive and negative sentiments and compare the accuracy of *Support Vector Machine* and *Naïve Bayes*. In this sentiment analysis using 2,339 datasets, the highest accuracy was achieved using the *Support Vector Machine* method at 91.67%, whereas the *Naïve Bayes* method only achieved 63.89%. Thus, it can be concluded that *Support Vector Machine* is superior in classifying sentiments regarding Pandawara Group's waste cleanup activities at Sukabumi Beach compared to *Naïve Bayes*.

Keywords: Sentiment Analysis; Trash Cleaning Activity; Social Media X; *Support Vector Machine*; *Naïve Bayes*

1. PENDAHULUAN

Era teknologi modern ini membuat internet semakin berkembang begitu pesat dengan mudahnya mengakses suatu internet bagi manusia dan perkembangan teknologi yang semakin canggih, [1] seperti contohnya teknologi *smartphone* dan jaringan internet yang sudah menuju sinyal 5G untuk memudahkan pengguna dalam mengakses suatu internet agar mendapatkan informasi dengan cepat dalam berbagai aspek seperti aspek kehidupan masyarakat, bisnis, pendidikan, dan masih banyak lainnya [2]. Hal ini membuat masyarakat lebih sering menggunakan media sosial yang bisa diakses oleh internet untuk mengakses sebuah informasi dan menjalin komunikasi dari jarak dekat maupun jarak jauh daripada menggunakan media cetak, seperti koran, majalah, atau media cetak lainnya yang dahulunya lebih sering digunakan [3].

Berkembangnya internet dan pengguna media sosial yang semakin banyak, serta kemudahan yang diberikan seperti salah satu contohnya yaitu media sosial X [4]. Pengguna internet pastinya sering menggunakan platform jejaring sosial, platform jejaring sosial yang banyak diakses pengguna internet salah satunya yaitu media sosial X, berdasarkan hasil riset *We Are School* tahun 2022, pengguna media sosial telah mencapai 191,4 juta jiwa dengan total keseluruhan populasi jiwa sebesar 277,7 juta jiwa [5]. Pada media sosial X, pengguna biasanya membagikan sebuah postingan atau membagikan sebuah pesan singkat terkait aktivitas sehari-hari antar pengguna ke pengguna lainnya yang disebut dengan *tweet* [6]. Oleh karena itu, *tweet* inilah yang akan digunakan untuk melihat bagaimana sentimen dari penggunaannya apakah menghasilkan opini-opini yang negatif atau positif mengenai isu tertentu yang terdapat pada media sosial X, selain itu media sosial X juga merupakan salah satu media sosial yang dapat mengutarakan pendapatnya secara bebas dengan opini tanpa batas yang dapat memudahkan para peneliti untuk melakukan analisis sentimen dan juga media sosial X merupakan sarana terbaik [7].

Pada aktivitas kegiatan sehari-hari manusia tidak luput dari sampah, hampir semua kegiatan yang dilakukan manusia akan menimbulkan sampah. Hingga saat ini, permasalahan global yang belum terpecahkan yaitu terkait sampah yang dihasilkan melalui aktivitas manusia [8]. Seiring berjalannya waktu dengan populasi yang akan semakin meningkat, jika tidak seimbang lahan pembuangan juga akan mempengaruhi sampah yang akan



semakin bertambah [9]. Kebiasaan buruk manusia yang diawali dengan membuang sampah sembarangan yang mengakibatkan pengaruh terhadap keseimbangan lingkungan, maka diperlukan langkah yang lebih mendalam dalam diberlakukannya kebijakan mengenai pengelolaan sampah [8].

Salah satu negara dengan populasi terpadat di dunia yaitu Indonesia dengan jumlah sampah meningkat pada awal tahun 2020 menjadi 32,8 juta ton/hari hingga tahun 2021 menjadi 41,2 juta ton/hari [10]. Sesuai hukum Pemerintahan Daerah dalam Undang-Undang Nomor 23 Tahun 2014, masalah tentang sampah termasuk tanggungjawab pemerintahan yang diserahkan ke daerah [11]. Namun, ada hal yang terjadi belakangan ini mengenai sebuah komunitas anak muda bernama Pandawara Grup yang ingin membantu membersihkan sampah di Pantai Loji tepatnya di daerah Sukabumi, namun aksi pembersihan sampah yang akan dilakukan tersebut ditolak oleh Kepala Desa dan Karang Taruna setempat dikarenakan Pandawara Grup mengatakan bahwa pantai tersebut merupakan pantai terkotor ke-4 di Indonesia dan tidak meminta izin terlebih dahulu yang membuat Kepala Desa Sangrawayang tidak terima atas pemberian predikat pantai terkotor ke-4 di Indonesia dan merasa bahwa pihak pemerintah seolah-olah tidak dilibatkan dalam kegiatan pembersihan sampah tersebut [12]. Kejadian tersebut yang viral, lalu menimbulkan banyak opini-opini sentimen masyarakat yang terdapat di *tweet* pada media sosial X untuk dilakukannya analisis sentimen untuk kepuasan terhadap suatu layanan [13].

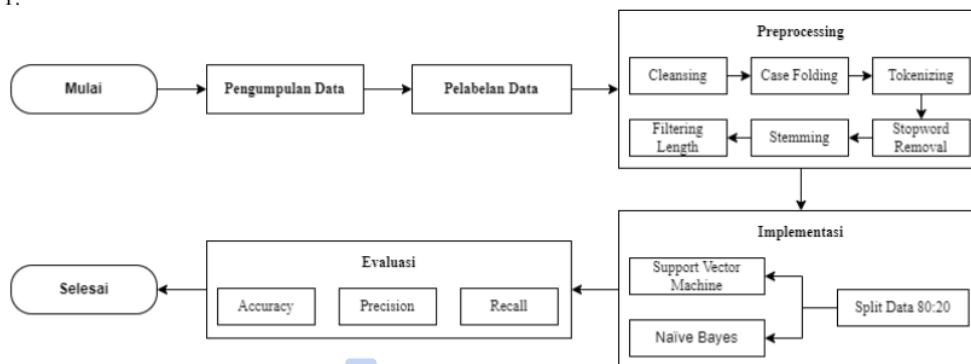
Analisis sentimen merupakan suatu teknik mendapatkan data komentar terhadap masyarakat untuk membantu menganalisis isi dari data yang didapatkan berupa pandangan ataupun opini negatif maupun positif terhadap suatu kejadian [14]. Data yang sudah didapatkan berupa komentar atau opini, kemudian dilakukan analisis sentimen untuk mengetahui sentimen perorangan yang dapat dijadikan tolak ukur untuk mengevaluasi suatu isu atau suatu layanan ke depannya [15].

23 Pada penelitian sebelumnya mengenai analisis sentimen terhadap pembobolan dan kebocoran data menggunakan metode *Naïve Bayes* dan *Support Vector Machine* dengan 1.083 *dataset* yang digunakan hingga menghasilkan akurasi 97,67% pada metode *Naïve Bayes* dan 73,75% pada metode *Support Vector Machine* [16]. Penelitian lainnya yaitu membandingkan metode *Support Vector Machine* dan *Naïve Bayes* terkait kebijakan penghapusan kewajiban skripsi dengan 1.303 *dataset* menghasilkan akurasi dari algoritma *Naïve Bayes* sebesar 75,00% dan *Support Vector Machine* sebesar 80,00% [17]. Penelitian lainnya mengenai penggunaan layanan BPJS dengan membandingkan ketiga algoritma yaitu KNN, *Decision Tree*, dan *Naïve Bayes* dengan data yang di *mining* sebanyak 903 *dataset*, lalu dari ketiga metode tersebut metode yang menghasilkan akurasi paling tinggi yaitu *Decision Tree* sebesar 96,13%, kedua KNN sebesar 95,58%, dan terakhir yaitu *Naïve Bayes* 89,14% [18]. Penelitian lainnya mengenai gadget berbagai aspek dengan membandingkan ketiga algoritma yaitu *Naïve Bayes*, SVM, dan k-NN dengan total 1.391 *dataset* hingga mendapatkan metode yang menghasilkan akurasi tertinggi yaitu metode SVM sebesar 96,43% [19]. Penelitian lainnya yaitu mengkomparasi metode *Support Vector Machine* dan *Naïve Bayes* dengan 2.334 *dataset* hingga menghasilkan akurasi paling tinggi pada algoritma *Support Vector Machine* sebesar 97,56% sedangkan *Naïve Bayes* sebesar 96,52% [20].

Dalam penelitian ini, peneliti memilih untuk membandingkan metode *Support Vector Machine* dan *Naïve Bayes* mengenai kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi, dikarenakan pada penelitian terdahulu belum pernah ada terkait analisis sentimen pada kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi. Tujuan dari penelitian ini yaitu, dari data yang sudah didapatkan, lalu dilakukan analisis sentimen untuk mengetahui sentimen negatif maupun positif, serta menganalisis perbandingan akurasi dari kedua metode *Support Vector Machine* dan *Naïve Bayes*.

2. METODOLOGI PENELITIAN

25 Tahap-tahap pada metode penelitian yang digunakan pada penelitian ini yaitu seperti yang terdapat pada gambar 1.



24 **Gambar 1.** Metodologi Penelitian



Tahapan pertama pada metodologi penelitian ini yaitu pengumpulan data, dimana data didapatkan melalui komentar *tweet* media sosial X yang di *crawling* menggunakan *tools* Google Colab, kemudian data yang sudah didapatkan dijadikan file berformat csv, selanjutnya data yang sudah berformat csv dilakukan pelabelan secara negatif maupun positif secara manual oleh peneliti, lalu data yang sudah berlabel sentimen dilanjutkan dalam tahapan *preprocessing*. Pada tahap *preprocessing* merupakan tahapan pembersihan data yang semula data berisi data yang sangat kotor menjadi data yang bersih dan lebih terstruktur agar dapat dengan mudah untuk diolah, pada tahapan ini berupa menghilangkan *mention*, hastag, link, spasi, simbol-simbol dan data duplikat, tidak hanya itu tetapi juga merubah semua sentimen menjadi huruf kecil, kemudian memisahkan sentimen menjadi kata per kata, lalu menghapus kata-kata yang tidak relevan atau tidak baku, selanjutnya kata-kata yang relevan atau baku dijadikan kata-kata dasar dengan menghilangkan imbuhan-imbuhannya, terakhir menghapus kata yang tidak terdiri dari 4 – 25 karakter. Sentimen yang sudah melalui tahap *preprocessing*, lalu dibagi menjadi 80% data uji dan 20% data latih yang kemudian dilanjutkannya tahapan pengimplementasian kedua metode yaitu metode *Support Vector Machine* dan *Naïve Bayes*, data yang sudah dilakukan tahapan pengujian metode, terakhir yaitu pengujian evaluasi untuk melihat dari sisi *accuracy*, *precision*, *recall* dari metode *Support Vector Machine* dan *Naïve Bayes*.

7

3. HASIL DAN PEMBAHASAN

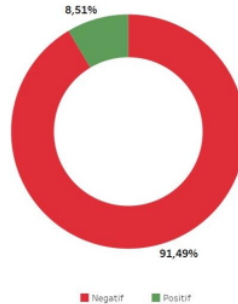
3.1 Pengumpulan Data

Tahap awal pada penelitian ini yaitu mengumpulkan data yang didapatkan melalui komentar *tweet* media sosial X dengan kata kunci pencarian “Pandawara Sukabumi” pada tanggal 6 Oktober 2023, kemudian data di *crawling* menggunakan bahasa pemrograman Python pada *tools* Google Colab, lalu *dataset* di *export* ke dalam file dengan format csv, seperti gambar 2 di bawah yang menunjukkan proses pengumpulan data.



Gambar 2. Proses Pengumpulan Data

3.2 Pelabelan Data



Gambar 3. Presentase Keseluruhan Sentimen

Data yang sudah didapatkan, kemudian tahap selanjutnya yaitu melabelkan *dataset* secara manual pada *tools* Microsoft Excel dengan label sentimen negatif dan positif. Pada proses pelabelan dihasilkan 2.876 sentimen negatif dan 251 sentimen positif seperti yang terdapat pada gambar 3 dari total keseluruhan 3.127 data, serta terdapat contoh dari beberapa data sentimen komentar *tweet* media sosial X yang terdapat pada tabel 1.

Tabel 1. Proses Pelabelan *Dataset*



Text	Sentimen
@dkristiawan_ malu kali tuh ketauan gitu....??? Ketauan selama ini si karang taruna ama Kadesnya kagak becus kerjanya Gimana nggak di cap jadi pantai terkotor kalo sampahnya udh sebanyak itu. Kalo gamau dicap jadi pantai terkotor ya bersihin dong pak! Giliran ada yg mau bantuin malah ga terima. Anehh banget kepala desa dan ketua karang tarunannyaðŸ™† Perkara pantainya diklaim jadi pantai terkotorâ€¦	Negatif
Alhamdulillah Pandawara berhasil menginspirasi dan meningkatkan kesadaran pemerintah. tidak perlu ajakan ekstrim cukup aksi nyata dan akan ada pihak2 yg tersentil olehnya Hebat sih pandawara grup berkat jasa mereka memviralkan pantai kotor tersebut jadi ada gerakan untuk membersihkannya. Segenap rakyat indonesia pastinya sangat berterimakasih kepada PG ðŸ™†?ðŸ™†?¼ https://t.co/gbz7n4Fcmt	Positif

3.3 Preprocessing

Data yang telah terlabeli, lalu dilakukan tahap selanjutnya yaitu *preprocessing*. Pada tahapan *preprocessing dataset* diolah yang semula data masih sangat kotor menjadi data yang bersih dan lebih terstruktur menggunakan *tools* RapidMiner untuk memudahkan dilakukan analisis sentimen. Tahap *preprocessing*, diantaranya yaitu:

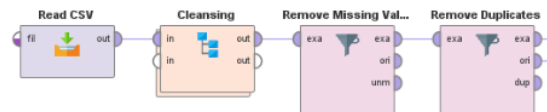
a. Cleansing



Gambar 4. Proses *Cleansing Dataset*

Pada proses *cleansing* terbagi lagi menjadi bagian-bagian seperti yang terdapat pada gambar 4, yaitu:

- 1. Replace URL**
Proses *cleansing* pertama yaitu *replace URL* untuk menghilangkan link-link yang terdapat pada komentar *tweet*.
- 2. Replace Simbol**
Proses *cleansing* kedua yaitu *replace simbol* untuk menghilangkan simbol-simbol yang terdapat pada komentar *tweet*.
- 3. Replace Mention**
Proses *cleansing* ketiga yaitu *replace mention* untuk menghilangkan mention atau tag kepada pengguna media sosial lainnya yang terdapat pada komentar *tweet*.
- 4. Replace Hastag**
Proses *cleansing* keempat yaitu *replace hastag* untuk menghilangkan kalimat bertagar yang terdapat pada komentar *tweet*.
- 5. Trim**
Proses *cleansing* kelima yaitu *trim* untuk menghilangkan spasi-spasi yang tidak berguna yang terdapat pada komentar *tweet*.



Gambar 5. Proses *Missing Values* dan *Remove Duplicates*

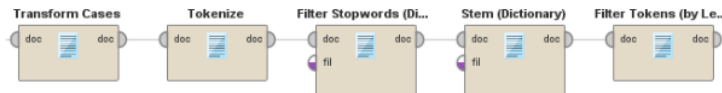
- 6. Remove Missing Values**
Proses *cleansing* keenam yaitu *remove missing values* untuk menghilangkan komentar *tweet* yang tidak memiliki nilai atau isi berupa komentar *tweet* di dalamnya.
- 7. Remove Duplicates**
Proses *cleansing* ketujuh yaitu *remove duplicates* untuk menghilangkan komentar *tweet* yang sama atau lebih dari satu tetapi memiliki isi yang sama pada komentar *tweet*.

Setelah dilakukan proses *cleansing* yang semula data masih sangat kotor menjadi data yang bersih dan lebih terstruktur agar lebih mudah untuk diolah analisis sentimennya, seperti contohnya yang terdapat pada tabel 2.

Tabel 2. Proses *Cleansing Dataset*



Sebelum	Sesudah
<p>@dkristiawan_ malu kali tuh ketauan gitu....???? Ketauan selama ini si karang taruna ama Kadesnya kagak becus kerjanya Gimana nggak di cap jadi pantai terkotor kalo sampahnya udh sebanyak itu. Kalo gamau dicap jadi pantai terkotor ya bersihin dong pak! Giliran ada yg mau bantuin malah ga terima. Anehh banget kepala desa dan ketua karang tarunannya Perkara pantainya diklaim jadi pantai terkotor Alhamdulillah Pandawara berhasil menginspirasi dan meningkatkan kesadaran pemerintah. tidak perlu ajakan ekstrim cukup aksi nyata dan akan ada pihak2 yg tersentil olehnya Hebat sih pandawara grup berkat jasa mereka memviralkan pantai kotor tersebut jadi ada gerakan untuk membersihkannya. Sege nap rakyat indonesia pastinya sangat berterimakasih kepada PG https://t.co/qbz7n4Fcmt</p>	<p>malu kali tuh ketauan gitu Ketauan selama ini si karang taruna ama Kadesnya kagak becus kerjanya Gimana nggak di cap jadi pantai terkotor kalo sampahnya udh sebanyak itu Kalo gamau dicap jadi pantai terkotor ya bersihin dong pak Giliran ada yg mau bantuin malah ga terima Anehh banget kepala desa dan ketua karang tarunannya Perkara pantainya diklaim jadi pantai terkotor Alhamdulillah Pandawara berhasil menginspirasi dan meningkatkan kesadaran pemerintah tidak perlu ajakan ekstrim cukup aksi nyata dan akan ada pihak yg tersentil olehnya Hebat sih pandawara grup berkat jasa mereka memviralkan pantai kotor tersebut jadi ada gerakan untuk membersihkannya Sege nap rakyat indonesia pastinya sangat berterimakasih kepada PG</p>



Gambar 6. Proses Case Folding sampai Filtering Length

b. Case Folding

Data yang sudah dilakukan proses *cleansing*, selanjutnya dilakukan proses *case folding*. *Case folding* merupakan proses mengubah seluruh huruf pada *dataset* menjadi huruf kecil menggunakan operator *transform cases* pada *tools* RapidMiner [21], seperti contohnya yang terdapat pada tabel 3.

Tabel 3. Proses Case Folding

Sebelum	Sesudah
<p>malu kali tuh ketauan gitu Ketauan selama ini si karang taruna ama Kadesnya kagak becus kerjanya Gimana nggak di cap jadi pantai terkotor kalo sampahnya udh sebanyak itu Kalo gamau dicap jadi pantai terkotor ya bersihin dong pak Giliran ada yg mau bantuin malah ga terima Anehh banget kepala desa dan ketua karang tarunannya Perkara pantainya diklaim jadi pantai terkotor Alhamdulillah Pandawara berhasil menginspirasi dan meningkatkan kesadaran pemerintah tidak perlu ajakan ekstrim cukup aksi nyata dan akan ada pihak yg tersentil olehnya Hebat sih pandawara grup berkat jasa mereka memviralkan pantai kotor tersebut jadi ada gerakan untuk membersihkannya Sege nap rakyat indonesia pastinya sangat berterimakasih kepada PG</p>	<p>malu kali tuh ketauan gitu ketauan selama ini si karang taruna ama kadesnya kagak becus kerjanya gimana nggak di cap jadi pantai terkotor kalo sampahnya udh sebanyak itu kalo gamau dicap jadi pantai terkotor ya bersihin dong pak giliran ada yg mau bantuin malah ga terima ane hh banget kepala desa dan ketua karang tarunannya perkara pantainya diklaim jadi pantai terkotor alhamdulillah pandawara berhasil menginspirasi dan meningkatkan kesadaran pemerintah tidak perlu ajakan ekstrim cukup aksi nyata dan akan ada pihak yg tersentil olehnya hebat sih pandawara grup berkat jasa mereka memviralkan pantai kotor tersebut jadi ada gerakan untuk membersihkannya sege nap rakyat indonesia pastinya sangat berterimakasih kepada pg</p>

c. Tokenizing

Proses selanjutnya yaitu *Tokenizing*, *Tokenizing* merupakan proses memisahkan komentar *tweet* yang seolah sebuah kalimat menjadi bagian kata per kata [22], seperti contohnya yang terdapat pada tabel 4.

Tabel 4. Proses Tokenizing



Sebelum	Sesudah
malu kali tuh ketauan gitu ketauan selama ini si karang taruna ama kadesnya kagak becus kerjanya	['malu', 'kali', 'tuh', 'ketauan', 'gitu', 'ketauan', 'selama', 'ini', 'si', 'karang', 'taruna', 'ama', 'kadesnya', 'kagak', 'becus', 'kerjanya']
gimana nggak di cap jadi pantai terkotor kalo sampahnya udh sebanyak itu kalo gamau dicap jadi pantai terkotor ya bersihin dong pak giliran ada yg mau bantuin malah ga terima anehh banget kepala desa dan ketua karang tarunannya perkara pantainya diklaim jadi pantai terkotor	['gimana', 'nggak', 'di', 'cap', 'jadi', 'pantai', 'terkotor', 'kalo', 'sampahnya', 'udh', 'sebanyak', 'itu', 'kalo', 'gamau', 'dicap', 'jadi', 'pantai', 'terkotor', 'ya', 'bersihin', 'dong', 'pak', 'giliran', 'ada', 'yg', 'mau', 'bantuin', 'malah', 'ga', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'dan', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantainya', 'diklaim', 'jadi', 'pantai', 'terkotor']
alhamdulillah pandawara berhasil menginspirasi dan meningkatkan kesadaran pemerintah tidak perlu ajakan ekstrim cukup aksi nyata dan akan ada pihak yg tersentil olehnya	['alhamdulillah', 'pandawara', 'berhasil', 'menginspirasi', 'dan', 'meningkatkan', 'kesadaran', 'pemerintah', 'tidak', 'perlu', 'ajakan', 'ekstrim', 'cukup', 'aksi', 'nyata', 'dan', 'akan', 'ada', 'pihak', 'yg', 'tersentil', 'olehnya']
hebat sih pandawara grup berkat jasa mereka memviralkan pantai kotor tersebut jadi ada gerakan untuk membersihkannya segenap rakyat indonesia pastinya sangat berterimakasih kepada pg	['hebat', 'sih', 'pandawara', 'grup', 'berkat', 'jasa', 'mereka', 'memviralkan', 'pantai', 'kotor', 'tersebut', 'jadi', 'ada', 'gerakan', 'untuk', 'membersihkannya', 'segenap', 'rakyat', 'indonesia', 'pastinya', 'sangat', 'berterimakasih', 'kepada', 'pg']

d. *Stopword Removal*

Sentimen yang sudah di *Tokenizing*, lalu dilakukan proses *Stopword Removal* yang merupakan proses penghapusan kata-kata yang tidak relevan atau tidak baku yang bereferensi dari file acuan pada website kaggle [23], seperti contohnya yang terdapat pada tabel 5.

Tabel 5. Proses *Stopword Removal*

Sebelum	Sesudah
['malu', 'kali', 'tuh', 'ketauan', 'gitu', 'ketauan', 'selama', 'ini', 'si', 'karang', 'taruna', 'ama', 'kadesnya', 'kagak', 'becus', 'kerjanya']	['malu', 'kali', 'tuh', 'ketauan', 'gitu', 'ketauan', 'si', 'karang', 'taruna', 'ama', 'kadesnya', 'kagak', 'becus', 'kerjanya']
['gimana', 'nggak', 'di', 'cap', 'jadi', 'pantai', 'terkotor', 'kalo', 'sampahnya', 'udh', 'sebanyak', 'itu', 'kalo', 'gamau', 'dicap', 'jadi', 'pantai', 'terkotor', 'ya', 'bersihin', 'dong', 'pak', 'giliran', 'ada', 'yg', 'mau', 'bantuin', 'malah', 'ga', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'dan', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantainya', 'diklaim', 'jadi', 'pantai', 'terkotor']	['gimana', 'nggak', 'cap', 'pantai', 'terkotor', 'kalo', 'sampahnya', 'udh', 'kalo', 'gamau', 'dicap', 'pantai', 'terkotor', 'ya', 'bersihin', 'giliran', 'yg', 'bantuin', 'ga', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantainya', 'diklaim', 'pantai', 'terkotor']
['alhamdulillah', 'pandawara', 'berhasil', 'menginspirasi', 'dan', 'meningkatkan', 'kesadaran', 'pemerintah', 'tidak', 'perlu', 'ajakan', 'ekstrim', 'cukup', 'aksi', 'nyata', 'dan', 'akan', 'ada', 'pihak', 'yg', 'tersentil', 'olehnya']	['alhamdulillah', 'pandawara', 'berhasil', 'menginspirasi', 'meningkatkan', 'kesadaran', 'pemerintah', 'ajakan', 'ekstrim', 'aksi', 'nyata', 'yg', 'tersentil']
['hebat', 'sih', 'pandawara', 'grup', 'berkat', 'jasa', 'mereka', 'memviralkan', 'pantai', 'kotor', 'tersebut', 'jadi', 'ada', 'gerakan', 'untuk', 'membersihkannya', 'segenap', 'rakyat', 'indonesia', 'pastinya', 'sangat', 'berterimakasih', 'kepada', 'pg']	['hebat', 'sih', 'pandawara', 'grup', 'berkat', 'jasa', 'memviralkan', 'pantai', 'kotor', 'gerakan', 'membersihkannya', 'segenap', 'rakyat', 'indonesia', 'pastinya', 'berterimakasih', 'pg']

e. *Stemming*

Data sentimen yang sudah tidak relevan, selanjutnya yaitu dilakukan proses *stemming*, dimana proses *stemming* merupakan menghapus kata-kata menjadi kata dasar dari kata-kata yang mengandung kata depan, kata sambung, kata ganti [24]. Proses *stemming* ini juga menggunakan referensi dari file acuan pada website gituhub [25], seperti contohnya yang terdapat pada tabel 6.

Tabel 6. Proses *Stemming*



Sebelum	Sesudah
['malu', 'kali', 'tuh', 'ketauan', 'gitu', 'ketauan', 'si', 'karang', 'taruna', 'ama', 'kadesnya', 'kagak', 'becus', 'kerjanya']	['malu', 'kali', 'tuh', 'tau', 'gitu', 'tau', 'si', 'karang', 'taruna', 'ama', 'kades', 'kagak', 'becus', 'kerja']
['gimana', 'nggak', 'cap', 'pantai', 'terkotor', 'kalo', 'sampahnya', 'udh', 'kalo', 'gamau', 'dicap', 'pantai', 'terkotor', 'ya', 'bersihin', 'giliran', 'yg', 'bantuin', 'ga', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantainya', 'diklaim', 'pantai', 'terkotor']	['gimana', 'nggak', 'pantai', 'kotor', 'kalo', 'sampah', 'udh', 'kalo', 'gamau', 'pantai', 'kotor', 'kotor', 'ya', 'bersihin', 'giliran', 'yg', 'bantuin', 'ga', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantai', 'klaim', 'pantai', 'kotor']
['alhamdulillah', 'pandawara', 'berhasil', 'menginspirasi', 'meningkatkan', 'kesadaran', 'pemerintah', 'ajakan', 'ekstrim', 'aksi', 'nyata', 'yg', 'tersentil']	['alhamdulillah', 'pandawara', 'hasil', 'inspirasi', 'tingkat', 'sadar', 'pemerintah', 'ajak', 'ekstrim', 'aksi', 'nyata', 'yg', 'sentil']
['hebat', 'sih', 'pandawara', 'grup', 'berkat', 'jasa', 'memviralkan', 'pantai', 'kotor', 'gerakan', 'membersihkannya', 'segenap', 'rakyat', 'indonesia', 'pastinya', 'berterimakasih', 'pg']	['hebat', 'sih', 'pandawara', 'grup', 'berkat', 'jasa', 'memviralkan', 'pantai', 'kotor', 'gerak', 'bersih', 'segenap', 'rakyat', 'indonesia', 'pasti', 'berterimakasih', 'pg']

f. *Filtering Length*

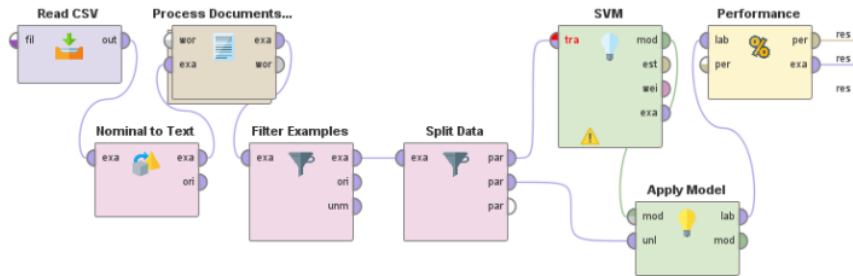
Proses terakhir dari tahapan *preprocessing* yaitu *filtering length*, pada proses ini kata yang tidak terdiri dari 4 sampai 25 huruf akan dihapus untuk memudahkan proses analisis sentimen selanjutnya [26], seperti contohnya yang terdapat pada tabel 7.

Tabel 7. Proses *Filtering Length*

Sebelum	Sesudah
['malu', 'kali', 'tuh', 'tau', 'gitu', 'tau', 'si', 'karang', 'taruna', 'ama', 'kades', 'kagak', 'becus', 'kerja']	['malu', 'kali', 'gitu', 'karang', 'taruna', 'kades', 'kagak', 'becus', 'kerja']
['gimana', 'nggak', 'cap', 'pantai', 'kotor', 'kalo', 'sampah', 'udh', 'kalo', 'gamau', 'cap', 'pantai', 'kotor', 'ya', 'bersihin', 'giliran', 'yg', 'bantuin', 'ga', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantai', 'klaim', 'pantai', 'kotor']	['gimana', 'nggak', 'pantai', 'kotor', 'kalo', 'sampah', 'kalo', 'gamau', 'pantai', 'kotor', 'bersihin', 'giliran', 'bantuin', 'terima', 'anehh', 'banget', 'kepala', 'desa', 'ketua', 'karang', 'tarunannya', 'perkara', 'pantai', 'klaim', 'pantai', 'kotor']
['alhamdulillah', 'pandawara', 'hasil', 'inspirasi', 'tingkat', 'sadar', 'pemerintah', 'ajak', 'ekstrim', 'aksi', 'nyata', 'yg', 'sentil']	['alhamdulillah', 'pandawara', 'hasil', 'inspirasi', 'tingkat', 'sadar', 'pemerintah', 'ajak', 'ekstrim', 'aksi', 'nyata', 'sentil']
['hebat', 'sih', 'pandawara', 'grup', 'berkat', 'jasa', 'memviralkan', 'pantai', 'kotor', 'gerak', 'bersih', 'segenap', 'rakyat', 'indonesia', 'pasti', 'berterimakasih', 'pg']	['hebat', 'pandawara', 'grup', 'berkat', 'jasa', 'memviralkan', 'pantai', 'kotor', 'gerak', 'bersih', 'segenap', 'rakyat', 'indonesia', 'pasti', 'berterimakasih']

3.4 Implementasi

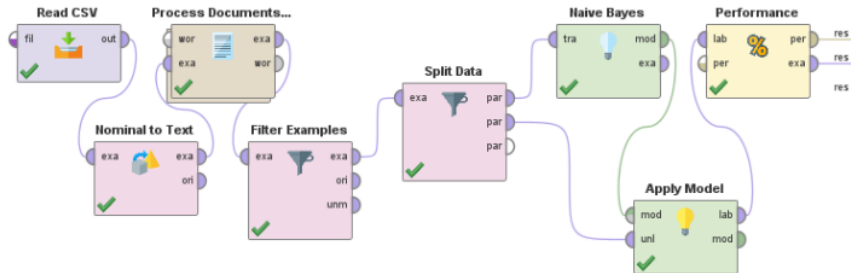
Data yang sudah melalui tahap *preprocessing*, selanjutnya yaitu tahapan implementasi. Tahapan implementasi pertama yaitu membagi *dataset* menjadi dua bagian, data uji dan data latih. Data uji dan data latih dibagi dengan persentase 80% untuk data uji dan 20% untuk data latih, semakin banyak data uji, semakin bagus juga metode akan mempelajari pola data untuk pengujian data latih untuk menghasilkan akurasi yang tinggi [27]. Total data yang sudah melalui tahap *preprocessing* hingga mendapatkan data sebanyak 2.339, dengan dibagi menjadi 1.871 data uji dan 468 data latih.





Gambar 7. Implementasi Metode *Support Vector Machine*

Data yang sudah dibagi menjadi data uji dan data latih, selanjutnya dilanjutkan pengimplementasian metode *Support Vector Machine* dengan melatih data sebanyak 468, setelah melatih data menggunakan metode *Support Vector Machine* nantinya akan menghasilkan *confusion matrix* yang berisi *accuracy*, *precision*, dan *recall* seperti contohnya proses implementasi dari metode *Support Vector Machine* yang terdapat pada gambar 7.



Gambar 8. Implementasi Metode *Naive Bayes*

Data yang sudah dibagi menjadi data uji dan data latih, selanjutnya dilanjutkan pengimplementasian metode *Naive Bayes* dengan melatih data sebanyak 468, setelah melatih data menggunakan metode *Naive Bayes* nantinya akan menghasilkan *confusion matrix* yang berisi *accuracy*, *precision*, dan *recall* seperti contohnya proses implementasi dari metode *Naive Bayes* yang terdapat pada gambar 8. Data yang sudah melalui proses implementasi metode *Support Vector Machine* dan *Naive Bayes*, selanjutnya yaitu melihat metode mana yang paling bagus dalam analisis sentimen ini dari hasil *confusion matrix* kedua metode yang digunakan.

3.5 Implementasi

Data yang sudah dilakukan implementasi kedua algoritma yaitu *Support Vector Machine* dan *Naive Bayes* seperti contohnya yang terdapat pada gambar 7 dan 8, kemudian akan menghasilkan hasil berupa *confusion matrix* yang berisi *accuracy*, *precision*, dan *recall* untuk melihat metode mana yang paling baik untuk digunakan dalam analisis sentimen terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi.

a. Evaluasi Metode *Support Vector Machine* dan *Naive Bayes*

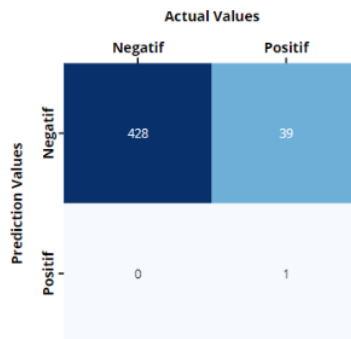
1. *Support Vector Machine*

Data yang sudah melalui tahapan implementasi metode *Support Vector Machine*, kemudian akan menghasilkan *confusion matrix*. Pada *confusion matrix* akan menghasilkan *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), dan *False Negatif* (FN) seperti yang terdapat pada tabel 8.

Tabel 8. Hasil *Confusion Matrix* Metode *Support Vector Machine*

	True Negatif	True Positif
Pred. Negatif	428	39
Pred. Positif	0	1

Berdasarkan implementasi dari metode *Support Vector Machine* yang terdapat pada gambar 9, menunjukkan bahwa *accuracy* yang dihasilkan sebesar 91,67%, *precision* sebesar 100%, dan *recall* sebesar 2,5%.



Gambar 9. Visualisasi *Confusion Matrix* Metode *Support Vector Machine*



$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} = \frac{(1+428)}{(1+428+0+39)} = \frac{429}{468} = 0,9167 \text{ atau } 91,67\% \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{1}{(1+0)} = \frac{1}{1} = 1 \text{ atau } 100\% \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{1}{(1+39)} = \frac{1}{40} = 0,025 \text{ atau } 2,5\% \quad (3)$$

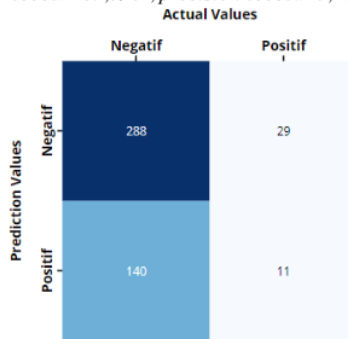
2. *Naïve Bayes*

Data yang sudah melalui tahapan implementasi metode *Naïve Bayes*, kemudian akan menghasilkan *confusion matrix*. Pada *confusion matrix* akan menghasilkan *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), dan *False Negatif* (FN) seperti yang terdapat pada tabel 9.

Tabel 9. Hasil *Confusion Matrix* Metode *Naïve Bayes*

	True Negatif	True Positif
Pred. Negatif	428	39
Pred. Positif	0	1

Berdasarkan implementasi dari metode *Naïve Bayes* yang terdapat pada gambar 10, menunjukkan bahwa *accuracy* yang dihasilkan sebesar 63,89%, *precision* sebesar 7,28%, dan *recall* sebesar 27,50%.



Gambar 10. Visualisasi *Confusion Matrix* Metode *Naïve Bayes*

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} = \frac{(11+288)}{(11+288+140+29)} = \frac{299}{468} = 0,6389 \text{ atau } 63,89\% \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{11}{(11+140)} = \frac{11}{151} = 0,07284 \text{ atau } 7,28\% \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{11}{(11+29)} = \frac{11}{40} = 0,275 \text{ atau } 27,5\% \quad (3)$$

b. Visualisasi *Wordcloud*

Visualisasi *wordcloud* merupakan memperlihatkan hasil kata yang paling sering muncul pada analisis sentimen yang telah dilakukan, semakin besar ukuran kata maka semakin sering juga kata tersebut muncul [28]. Pada keseluruhan *dataset* sebesar 2.339 data dihasilkan visualisasi *wordcloud* pada data bersentimen positif berwarna hijau sedangkan data bersentimen negatif berwarna merah, seperti yang terdapat pada gambar 11 dan gambar 12.



Gambar 11. *Wordcloud* Sentimen Positif



Berdasarkan *wordcloud* yang terdapat pada gambar 11, memperlihatkan kata yang paling sering muncul pada sentimen positif yaitu kata “pandawara”, “pantai”, dan “bisa”.



Gambar 12. *Wordcloud* Sentimen Positif

Berdasarkan *wordcloud* yang terdapat pada gambar 12, memperlihatkan kata yang paling sering muncul pada sentimen negatif yaitu berupa bentuk kekecewaan masyarakat terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi namun dibatalkan oleh kepala desa dan karang taruna setempat, seperti kata “sampah”, “kades”, dan “bersihin”.

4. KESIMPULAN

Berdasarkan analisis sentimen yang dilakukan terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi, dihasilkan *dataset* akhir yang mampu diproses hingga akhir ke dalam implementasi metode *Support Vector Machine* dan *Naïve Bayes* sebanyak 2.339 data, dengan 199 data bersentimen positif dan 2.140 bersentimen negatif yang berisi banyak kekecewaan terhadap kebijakan pemerintahan yang berada di wilayah Sukabumi terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi. Hasil dari pengujian metode *Support Vector Machine* dan *Naïve Bayes* menghasilkan *accuracy* sebesar 91,67% pada metode *Support Vector Machine* dan 63,89% pada metode *Naïve Bayes*, mengartikan bahwa *accuracy* metode *Support Vector Machine* lebih tinggi dan lebih baik untuk mengklasifikasikan sentimen terhadap kegiatan pembersihan sampah yang dilakukan Pandawara Grup di pantai Sukabumi dibandingkan metode *Naïve Bayes*. Banyaknya sentimen negatif dibandingkan sentimen positif pada analisis sentimen ini, seharusnya dapat menjadi bahan pertimbangan bagi pemerintahan daerah setempat dalam mengambil keputusan mengenai kebijakan-kebijakan agar tidak mengecewakan masyarakat terkait suatu layanan dan dapat meningkatkan kinerja pemerintahan ke depannya. Rekomendasi saran dari peneliti untuk penelitian selanjutnya, diharapkan peneliti selanjutnya dapat lebih memperhatikan terhadap sentimen pendapat pengguna pada tahapan pelabelan *dataset* agar pada sentimen label positif dan negatif mendapatkan hasil yang lebih baik dan akurat.

REFERENCES

Dendy Aprilianto Nugroho - Analisis Sentimen Terhadap Kegiatan Pembersihan Sampah Pada Media Sosial X Menggunakan Metode Support Vector Machine dan Naïve Bayes

ORIGINALITY REPORT

11%

SIMILARITY INDEX

9%

INTERNET SOURCES

7%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1	prosiding.unimus.ac.id Internet Source	1%
2	ejurnal.methodist.ac.id Internet Source	1%
3	Submitted to Tarumanagara University Student Paper	1%
4	djournals.com Internet Source	1%
5	Submitted to Universitas Brawijaya Student Paper	<1%
6	eksplora.stikom-bali.ac.id Internet Source	<1%
7	repository.uin-suska.ac.id Internet Source	<1%
8	repository.unj.ac.id Internet Source	<1%

9

123dok.com

Internet Source

<1 %

10

Sopian Aji, Ina Maryani, Elly Muningsih. "Analisis Sentiment Masyarakat Menggunakan Penggabungan Algoritma Naive Bayes Dan Particle Swarm Optimization", IJCIT (Indonesian Journal on Computer and Information Technology), 2022

Publication

<1 %

11

lib.unnes.ac.id

Internet Source

<1 %

12

Antonius Mbay Ndapamuri, Danny Manongga, Ade Iriani. "Analisis Sentimen Ulasan Aplikasi Tripadvisor Dengan Metode Support Vector Machine, K-Nearest Neighbor, Dan Naive Bayes", INOVTEK Polbeng - Seri Informatika, 2023

Publication

<1 %

13

Herlawati Herlawati, Rahmadya Trias Handayanto, Prima Dina Atika, Fata Nidaul Khasanah et al. "Analisis Sentimen Pada Situs Google Review dengan Naive Bayes dan Support Vector Machine", Jurnal Komtika (Komputasi dan Informatika), 2021

Publication

<1 %

14

ejournal.unesa.ac.id

Internet Source

<1 %

15 Doni Winarso, Yanda Noor Yudha, Syahril.
"Analisis Sentimen Masyarakat Pada Twiter
Terhadap Isu Covid-19 Menggunakan Metode
Lexicon Based", JURNAL FASILKOM, 2021

Publication

<1 %

16 Nurlaelatul Maulidah, Riki Supriyadi, Dwi Yuni
Utami, Fuad Nur Hasan, Ahmad Fauzi, Ade
Christian. "Prediksi Penyakit Diabetes Melitus
Menggunakan Metode Support Vector
Machine dan Naive Bayes", Indonesian
Journal on Software Engineering (IJSE), 2021

Publication

<1 %

17 Submitted to Sriwijaya University

Student Paper

<1 %

18 repository.lppm.unila.ac.id

Internet Source

<1 %

19 Submitted to Dewan Perwakilan Rakyat

Student Paper

<1 %

20 www.neliti.com

Internet Source

<1 %

21 eprints.undip.ac.id

Internet Source

<1 %

22 jurnal.polibatam.ac.id

Internet Source

<1 %

23 jurnalti.polinema.ac.id

Internet Source

<1 %

24	jutif.if.unsoed.ac.id Internet Source	<1 %
25	www.grafiati.com Internet Source	<1 %
26	www.ifxid.com Internet Source	<1 %
27	Fajar Sidik, Ibnu Suhada, Azhar Haikal Anwar, Firman Noor Hasan. "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier", Jurnal Linguistik Komputasional (JLK), 2022 Publication	<1 %
28	Fitri Adha Hariyati Airi, Tati Suprapti, Agus Bahtiar. "KOMPARASI METODE KLASIFIKASI DATA MINING UNTUK PREDIKSI PENYAKIT STROKE", E-Link: Jurnal Teknik Elektro dan Informatika, 2023 Publication	<1 %
29	id.123dok.com Internet Source	<1 %
30	journal.umsida.ac.id Internet Source	<1 %
31	judi.port5.com Internet Source	<1 %
32	www.blogtekno.us Internet Source	<1 %

33

Jimmy, Eni Heni Hermaliani, Laela Kurniawati.
"ANALISIS KLASIFIKASI SENTIMEN
PENGGUNA MEDIA SOSIAL TWITTER
TERHADAP PENUNDAAN PEMILU PRESIDEN
TAHUN 2024", Jurnal Indonesia : Manajemen
Informatika dan Komunikasi, 2023

Publication

<1 %

34

Risawati Risawati, Siti Ernawati, Ina Maryani.
"OPTIMASI PARAMETER PSO BERBASIS SVM
UNTUK ANALISIS SENTIMEN REVIEW JASA
MASKAPAI PENERBANGAN BERBAHASA
INGGRIS", EVOLUSI : Jurnal Sains dan
Manajemen, 2020

Publication

<1 %

35

konsultasiskripsi.com

Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off