

ANALISIS SENTIMEN KOMENTAR TIKTOK TERKAIT WAFATNYA LULA LAHFH MENGUNAKAN ALGORITMA SVM

Sentiment Analysis of Tiktok Comments on Lula lahfah's Death SVM

Huana Novalin Sineri¹⁾, Florenta Demanovensa Lewataka²⁾, Christian Yulianto Thiong Manibuy³⁾.

^{1,2,3} Teknik Informatika, Fakultas Teknik, Universitas Papua
Jl. Gunung Salju Amban, Manokwari, Papua Barat, Indonesia
E-mail : hanavalin770@gmail.com¹⁾, florentademano@gmail.com²⁾,
christianmanibuy744@gmail.com³⁾.

Abstract

The objective of this investigation is to evaluate the sentiment within Tiktok user komentar regarding the wafat of social media personality Lula lahfah by applying the Support Vector Machine (SVM) algoritma. The material was collected through scraping techniques from three Tiktok handles: @detikcom, @liputan6com, and @antaranews, yielding a total of 10,701 komentar. Once repetitive records were excluded, 5,472 polished observations were retrieved, yang then proceeded through preprocessing phases such as cleaning, case folding, word normalization, tokenizing, stopword removal, and stemming. Afterward, the records were mechanically categorized via a lexicon approach and split into training data (80%) and test data (20%). This classification task was conducted through the Support Vector Machine (SVM) algoritma integrated with TF-IDF feature extraction. The outcomes indicated yang the SVM architecture operated optimally, reaching an accuracy level of 98% with a 2% error rate. Additionally, the precision, recall, and F1-score statistics for each sentiment class (positif, netral, and negatif) fluctuated from 0.96 to 1.00. Consequently, the SVM algoritma has exhibited high competence in distinguishing sentiment in social media comment data and is capable of providing an accurate picture of public opinion trends toward an event.

Keywords: Sentiment Analysis, Tiktok, Support Vector Machine, TF-IDF, Text Mining

Abstrak

Riset ini bermaksud guna menelaah sentimen komentar pengakses Tiktok berkenaan fenomena wafat pesohor internet *Lula Lahfah* memakai algoritma *Support Vector Machine (SVM)*. Informasi dikumpulkan lewat metode *scraping* dari tiga kanal Tiktok, yakni @detikcom, @liputan6com, serta @antaranews, berjumlah keseluruhan 10.701 komentar. Usai dilakukan eliminasi catatan ganda, didapatkan 5.472 dokumentasi murni yang lantas diolah melewati tingkatan *preprocessing* yang mencakup *cleaning*, *case folding*, standarisasi istilah, *tokenizing*, *stopword removal*, beserta *stemming*. Berikutnya, materi dikategorisasi secara mekanis mengaplikasikan perspektif *lexicon* serta dipisahkan menjadi *data training* (80%) maupun *data testing* (20%). Prosedur pengelompokan dilangsungkan memakai algoritma *Support Vector Machine (SVM)* lewat *feature extraction TF-IDF*. Temuan observasi memperlihatkan jika kerangka kerja *SVM* mempunyai kapabilitas yang amat unggul lewat angka ketepatan senilai 98% dan kekeliruan sebesar 2%. Disamping itu, besaran *precision*, *recall*, serta *F1-score* bagi tiap kategori sentimen (positif, netral, dan negatif) terletak di cakupan 0,96 sampai 1,00. Oleh karena itu, algoritma *SVM* teruji berdaya guna dalam mengerjakan pemilahan sentimen terhadap materi komentar platform daring serta sanggup menyajikan visualisasi yang tepat mengenai orientasi pandangan publik terhadap suatu peristiwa.

Kata Kunci: Analisis Sentimen, Tiktok, Support Vector Machine, TF-IDF, Text Mining

1. Pendahuluan

Transformasi sarana digital sudah mentransformasi gaya publik kala mengutarakan aspirasi dan merespons berbagai peristiwa secara cepat. Platform digital seperti TikTok memungkinkan pengguna berpartisipasi melalui fitur komentar, sehingga opini

publik dapat muncul dalam waktu singkat dan berpotensi membentuk persepsi kolektif terhadap suatu isu yang sedang viral.[1] karena komentar yang ditulis pengguna bersifat spontan dan beragam, pola sentimen di dalamnya sering kali sulit dipahami secara manual ketika volume percakapan meningkat.

Salah satu peristiwa yang menarik perhatian publik di media sosial adalah meninggalnya selebgram Lula Lahlah. Komentar-komentar pada unggahan terkait peristiwa tersebut menunjukkan berbagai sentimen, mulai dari ungkapan duka cita, empati, dukungan, hingga respons kritis dan bentuk ketidakpercayaan terhadap informasi yang beredar. Keragaman ini menjadi komentar Tiktok sebagai data yang relevan untuk dianalisis secara lebih sistematis guna mengetahui kecenderungan opini publik secara keseluruhan (Fikri & Setiawan, 2023).

Namun, analisis komentar dalam skala besar menghadapi dua tantangan utama. Pertama, pengumpulan data komentar secara manual kurang efisien dan rentan terhadap subjektivitas. Oleh karena itu, [2] dalam penelitian ini proses pengumpulan data dilangsungkan secara mandiri lewat metode *scraping* memakai bahasa kode Python pada Google Collab untuk mendapatkan komentar dalam jumlah besar secara cepat dan terstruktur. kedua, teks komentar Tiktok umumnya tidak terstruktur, mengandung bahasa tidak baku, singkatan, emoji, serta noise berupa simbol dan tautan yang dapat mengganggu analisis. Maka, diperlukan tahapan **pra-pemrosesan teks** guna memurnikan dokumentasi supaya layak guna dalam tahapan (Saputra & Pratama, 2023). [3]

Analisis sentimen ialah komponen *text mining* yang berorientasi mendeteksi polaritas afeksi maupun pandangan pada sebuah naskah terhadap subjek spesifik. Pada riset ini, analisis sentimen difokuskan pada pengguna Tiktok terkait meninggalnya Lula Lahlah dan diklasifikasikan ke dalam kategori **positif, netral, dan negatif** (Syahputra et al., 2023). [4] klasifikasi ini diharapkan mampu memberikan

klasifikasi ini klasifikasi ini diharap bisa menyuguhkan deskripsi yang lebih faktual perihal orientasi sentimen massa terhadap peristiwa tersebut (Prameswari & Kurniawan, 2022). Untuk menjalankan prosedur pemilahan sentimen, kajian ini memakai algoritma **Support Vector Machine (SVM)**. [5] SVM ialah salah satu teknik *machine learning* yang ampuh bagi pengategorian narasi lantaran yang efektif untuk klasifikasi teks karena mampu membangun hyperplane pemisah yang optimal antar kelas dalam ruang fitur. Kelebihan utama SVM terutama berlokasi di kapabilitasnya menangani data representasi teks hasil ekstraksi fitur (Joachims, 1998). [6] Dengan demikian, SVM dipilih sebagai metode yang sesuai untuk pemodelan sentimen komentar Tiktok.

Kinerja model SVM dipengaruhi oleh beberapa faktor, seperti kualitas data, hasil pra-pemrosesan, teknik ekstraksi fitur, serta pemilihan parameter model (Cortes & Vapnik, 1995). [3] Maka dari itu, pengamatan ini melakukan pengujian kerangka secara kuantitatif memakai **metrik akurasi, presisi, recall, serta F1-score** supaya kompetensi pemilahansentimen dapat dinilai secara komprehensif dan tidak hanya berdasarkan satu indikator.

Berdasarkan uraian tersebut, riset ini berupaya menelaah sentimen komentar Tiktok bertalian peristiwa wafat tokoh media sosial *Lula Lahlah* menerapkan algoritma *Support Vector Machine (SVM)*. [6] Temuan analisis diproyeksikan bisa menyediakan pemahaman perihal arah aspirasi khalayak berbarengan membuktikan keberhasilan implementasi SVM dalam kategorisasi sentimen pada dokumentasi komentar jejaring sosial

2.2 Text Mining

Text Mining ialah prosedur pemrosesan data guna meraih wawasan yang berdaya guna melewati pelbagai fase layaknya *cleaning, case folding, tokenizing, stopword removal*, serta *stemming*. [2] Pada analisis sentimen, text mining juga berperan penting karena sebagian besar data komentar media sosial bersifat tidak terstruktur dan mengandung bahasa tidak baku. [4] Oleh karena itu, preprocessing ini menjadi tahapan krusial guna meningkatkan mutu hasil klasifikasi.

2. Tinjauan Pustaka

2.1 Analisis Sentimen

Analisis Sentimen ialah elemen *text mining* yang dimanfaatkan buat mengenali, mengolah, serta mengelompokkan aspirasi atau perasaan di dalam tulisan ke kategori tertentu semisal positif, netral, serta negatif. Analisis sentimen ini banyak diterapkan pada media sosial karena pengguna sering menyampaikan pendapat secara langsung melalui komentar atau unggahan. [1]

Dengan adanya analisis sentimen ini, informasi mengenai kecenderungan opini publik dapat diperoleh secara lebih cepat dan sistematis.

Maka dari itu, *preprocessing* ialah jenjang krusial guna meningkatkan kualitas hasil klasifikasi.

2.3 Support Vector Machine (SVM) Support Vector

Machine (SVM) ialah satu di antara algoritma *machine learning supervised* yang dioperasikan buat prosedur pemilahan informasi.[5] *SVM* pun beroperasi lewat pencarian *hyperplane* paling unggul yang sanggup memecah dokumentasi ke kategori-kategori spesifik lewat *margin* yang maksimal. Teknik tersebut diakui berdaya guna pada pengelolaan data bervolume kompleks layaknya materi naskah keluaran *feature extraction TF-IDF*. [6]

Dalam Penelitian ini analisis sentimen, algoritma SVM sering dipilih lantaran mempunyai derajat ketepatan yang amat mumpuni yang sanggup menciptakan hasil pengelompokan yang paling prima terhadap dokumentasi tulisan jejaring daring.[7]

2.4 Penelitian Terdahulu

Penelitian ini mengenai analisis sentimen di platform daring sudah kerap dilangsungkan memakai bermacam metode klasifikasi. Salah satu penelitian dilakukan oleh peneliti Jurnal “**Analisis Sentimen Aplikasi Tiktok Menggunakan Algoritma Naive Bayes dan Support Vector Machine**” yang mengomparasikan kinerja

algoritma *Naive Bayes* dan *SVM* kala memilah sentimen pemakai Tiktok (Sebastiani, 2002).[8] Temuan riset memaparkan jika algoritma *SVM* mempunyai hasil pengategorian yang kian unggul daripada metode lainnya.

Penelitian lain mengenai analisis sentimen komentar TikTok pada video viral menggunakan metode *Naive Bayes* menunjukkan bahwa *preprocessing* dan ekstraksi fitur memiliki pengaruh besar terhadap tingkat akurasi model klasifikasi (Fitriani & Nugroho, 2023).[8]

Berdasarkan penelitian terdahulu tersebut, riset ini mengaplikasikan algoritma *Support Vector Machine (SVM)* guna menelaah sentimen komentar TikTok mengenai meninggalnya selebgram Lula Lahlah guna mengetahui kecenderungan opini publik terhadap peristiwa tersebut.[9].

3. Metodologi Penelitian

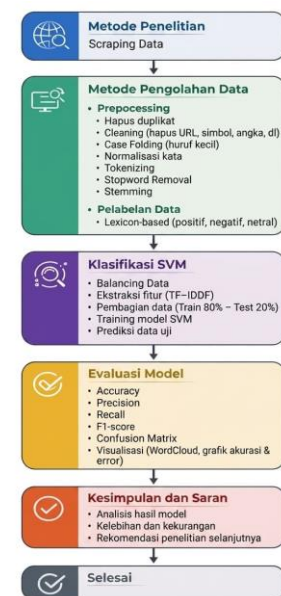
Dalam observasi ini dioperasikan algoritma *Support Vector Machine (SVM)* buat mengelompokkan sentimen komentar pemakai Tiktok terkait peristiwa meninggalnya Lula Lahlah. Proses penelitian dilakukan melalui beberapa tahap, diawali dengan akuisisi informasi, diteruskan lewat pemrosesan materi, implementasi algoritma *Support Vector Machine (SVM)*, sampai penaksiran temuan.[6]

3.1 Pengumpulan data

Penetapan teknik ini dilandasi eskalasi pemanfaatan wadah digital selaku instrumen kala mengemukakan pandangan khalayak berkenaan sebuah kejadian. Tiktok selaku satu di antara wadah daring yang tersohor memberi peluang pengakses guna bersosialisasi serta menyalurkan respons dengan langsung melalui komentar.[10]

Peristiwa meninggalnya Lula Lahlah menjadi topik yang banyak diperbincangkan di Tiktok, sehingga menghasilkan berbagai opini dari masyarakat yang dapat dianalisis menggunakan teknik analisis sentimen (Wibowo & Firmansyah, 2022).[12] Data pada riset ini didapatkan lewat komentar pemakai Tiktok yang diambil dari beberapa video yang membahas peristiwa tersebut (Nurhayati, 2021).

Prosedur perolehan informasi dilangsungkan memakai metode *scraping (crawling)* materi mengoperasikan Tiktok *Comment Scraper* asal *apify*



Gambar 1 Urutan Penelitian

Dokumentasi yang sukses dihimpun lantas diarsipkan pada bentuk .csv atau .xlsx buat kebutuhan penelitian lebih mendalam(Liu,2012). [16]

3.2 Pengolahan Data

Di fase ini, data diproses melewati fase *pre-processing* beserta pengategorian informasi. *Pre-processing* ialah jenjang awal dalam pengolahan data text yang berfungsi untuk meningkatkan kualitas data dengan cara membersihkan, menyederhanakan, dan menstandarisasi teks sehingga dapat diproses secara optimal oleh algoritma(Feldman & Sanger, 2007). [9]

Tujuannya ialah guna mengeskali kualitas informasi, meminimalkan distraksi (*noise*), membenahi pertentangan, menyelaraskan ukuran serta meniadakan fragmen-fragmen yang tidak bersangkutan. Prosedur tersebut menjamin bahwasanya materi yang dipakai di edukasi model mempunyai standar yang prima,[2] hingga menciptakan estimasi yang presisi.

Pemberian identitas informasi dilangsungkan buat menyematkan tanda pada tiap data, supaya bisa dikenali apakah muatannya menyertakan sentimen positif, netral, atau negatif. Operasi tersebut dieksekusi secara mekanis lewat penggunaan kamus *lexicon*, yang berisi himpunan istilah beserta angka bebannya buat menolong menetapkan sentimen di naskah. [15]

Skor polaritas dikalkulasi bersandar pada jumlah beban istilah di tiap komentar, serta temuannya diaplikasikan selaku tanda guna memastikan apakah komentar itu mengandung sentimen positif, netral, atau negatif(Han et al., 2012).[11]

Komentar diklasifikasikan positif jika skor polaritasnya melampaui angka nihil, netral andai skor polaritasnya ekuivalen angka nihil, serta negatif tatkala skor polaritasnya di bawah angka nol(Kurniawan & Setiawan, 2022).[17]

3.3 Klasifikasi Support Vector Machine

Algoritma Support Vector Machine (SVM) ialah algoritma *supervised* yang dimanfaatkan buat mengerjakan pemilahan materi lewat strategi mencari sekat pemecah paling ideal (*hyperplane*) yang sanggup memilah dokumentasi ke kategori-kategori spesifik(Goldberg & Levy, 2014).[5] SVM

beroperasi lewat optimalisasi *margin* di antara informasi dari kategori berbeda hingga melahirkan kerangka lewat derajat generalisasi yang unggul. Di pengamatan tersebut, algoritma SVM dipakai guna mengelompokkan sentimen komentar Tiktok pada klasifikasi positif, netral, serta negatif berlandaskan atribut yang telah diambil memakai teknik *TF-IDF* (*Term Frequency-Inverse Document Frequency*)(Suartika & Arsi, 2022).[6] Nilai lebih SVM dalam mengelola materi berukuran kompleks menjadikannya teknik yang ampuh pada penelaahan sentimen berbasis naskah

3.4 Evaluasi

Evaluasi model pada riset dilangsungkan guna mengestimasi performa algoritma *Support Vector Machine* (SVM) dalam memilah sentimen komentar Tiktok ke klasifikasi positif, netral, serta negatif.[12] Prosedur penaksiran memakai bermacam ukuran, yakni *confusion matrix*, ketepatan, kekeliruan, beserta *classification report* yang melingkupi *precision*, *recall*, dan *f1-score*. *Confusion matrix* dioperasikan guna memahami kuantitas estimasi yang tepat maupun keliru di tiap-tiap kategori, sehingga dapat memberikan gambaran distribusi hasil klasifikasi secara lebih rinci. [18] Selain itu, akurasi digunakan untuk mengukur tingkat ketepatan model secara keseluruhan, sedangkan error menunjukkan tingkat kesalahan prediksi yang dihasilkan oleh model(Raschka & Mirjalili, 2019).[16] *Classification report* digunakan untuk mengevaluasi kinerja kerangka secara kian mendetail lewat angka *precision*, *recall*, dan *f1-score* di tiap kategori sentimen. Lewat pengoperasian perpaduan matriks penaksiran itu, daya guna model bisa ditaksir secara menyeluruh serta tidak hanya bergantung pada satu indikator saja(Pak & Paroubek, 2010).[10]

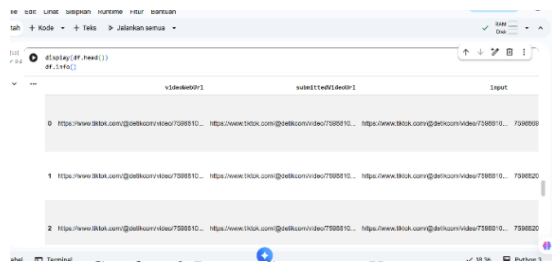
4. Hasil Dan Pembahasan

4.1 Pengumpulan Data.

4.1.1 Scraping Data Tiktok

Prosedur perolehan informasi dilangsungkan terhadap 3 tayangan yang berlainan dari sejumlah akun yang berbeda, yakni dari akun @detikcom, @liputan6com, @antaranews.

Huana N Sineri., Dkk., Analisis Sentimen Komentar Tiktok Terkait Wafatnya Lula lahfah Menggunakan Algoritma SVM



Gambar 2 Proses Scrapping Komentar

Pada gambar 2 menunjukan adanya prosedur *scraping* komentar tiktok di wadah *web scraping apify.com* dari 3 tayangan yang berlainan di kanal @detik.com, @liputan6com, dan @antaranews. Berikutnya seluruh komentar yang telah di-*scraping* diarsipkan dalam format .csv, setelahnya dikonversi ke bentuk .xlsx.



Gambar 3 Hasil Scrapping komentar

Gambar 3 ini menunjukan temuan dari *scraping* komentar pada unggahan di Tiktok kanal @detik.com, @liputan6com, dan @antaranews. Yang menghasilkan materi 10701 komentar, tayangan pada kanal @detik.com menghasilkan 4149 komentar, tayangan kanal @liputan6com menghasilkan 3792 komentar, dan tayangan pada kanal @antaranews menghasilkan 2760 komentar. Tabel 1 memaparkan visualisasi temuan *scraping* komentar dalam wujud tabel.

Tabel 1 Hasil Scrapping Komentar Tiktok

Akun tiktok	Tema video	jumlah data
@detik.com	Tangis Reza Arap Pecah Usai Pemakaman Lula Lahfah	4149 Komentar.
@liputan6com	Ada Bekas Bercak Darah Di Apartemen Lula Lahfah	3792 Komentar.
@antaranews	Kabar Duka, Selebgram Lula Lahfah Kekasih Reza Arap Meninggal Dunia Di Usia 26 Tahun.	2760 Komentar
Total		10701 Komentar

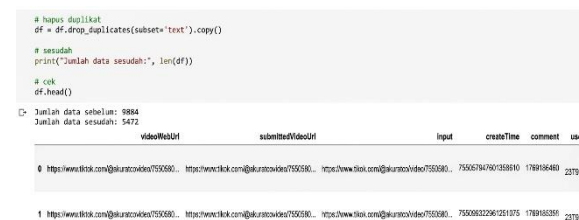
4.2 Pre-processing Data

Tahapan pra-pemrosesan (pre-processing)

merupakan proses menganalisis dan mengekstrak informasi yang berguna dengan melibatkan *text mining*. Dengan implementasi *text mining*, suatu prosedur repetitif dua jenjang diawali dengan pengorganisasian tekstual materi mentah, yang lantas didayagunakan guna menggali inteligensi serta pemahaman yang bersesuaian. Berikut ialah fase-fase *pre-processing* data:

4.2.1 Hapus Data Duplikat

Guna menjamin standar serta keutuhan informasi telaah, materi ganda dieliminasi. Komentar yang dinilai ganda ialah komentar yang naskahnya benar-benar serupa. Prosedur ini bermaksud buat mengenali data yang sama (ganda) guna menangkal redundansi serta bias, khususnya dalam analisis sentimen yang memerlukan materi representatif dan unik. Sasarannya buat menjauhi penyimpangan temuan telaah akibat komentar yang berulang, sehingga menciptakan penelaahan yang lebih valid dan objektif.



Gambar 4 Hapus Data Duplikat

Gambar 4 menunjukan yakni telah terjadi duplikasi berlebihan pada data atau *scraping* sehingga sebanyak 4,332 dihapus dari 9,804 data awal dan tersisa 5,472 data bersih yang siap di proses lebih lanjut. Duplikasi berlebihan biasanya disebabkan oleh proses *scraping* yang berulang sehingga terjadi penumpukan atau komentar yang sama persis tapi dari pengguna yang berbeda. Banyak dari pengguna menuliskan komentar yang sama persis, misalnya “Turut berduka cita”, “amin”, “RIP”, dan lain sebagainya.

4.2.2 Cleaning Data

Penjernihan data (*data cleaning*) ialah tingkatan di dalam pra-pemrosesan informasi yang berorientasi guna mengeskalasi mutu *dataset*

melalui pendeteksian serta penanggulangan angka yang lenyap (*missing values*), informasi ganda, kekeliruan entri, beserta ketidakselarasan bentuk. Prosedur tersebut krusial guna menjamin keabsahan serta keandalan temuan analisis yang dilakukan pada tahap selanjutnya.

index	text	clean_text
0	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: 'Kataanya ngabalin? lu apaan makudnya?'	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: Kataanya ngabalin? lu apaan makudnya?
1	haahh	haahh
2	salit apa lurnd berdukaca	salit apa lurnd berdukaca
3	innallahwainnallahroju ya allah	innallahwainnallahroju ya allah
4	tadi baca koment katanya lagi diel yaallah	tadi baca koment katanya lagi diel yaallah
5	please gals jangan mengging opini	please gals jangan mengging opini
6	fy high laura lula	fy high laura lula
7	yg pin respon di lid	yg pin respon di lid
8	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga

Gambar 5 Cleaning Data

Gambar 5 menyajikan temuan penjernihan materi yang mencakup eliminasi *retweet*, simbol ortografis, *mention* (@), *URL*, bilangan yang tidak bersangkutan, beserta aksara istimewa lainnya yang tidak dipergunakan buat penelaahan kian mendalam.

4.2.3 Case Folding

Case folding ialah metode alih bentuk naskah yang mengubah seluruh karakter menjadi huruf kecil, dengan tujuan mempermudah identifikasi dan pemrosesan informasi.

index	text	casefold_text
0	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: 'Kataanya ngabalin? lu apaan makudnya?'	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: Kataanya ngabalin? lu apaan makudnya?
1	haahh	haahh
2	salit apa lurnd berdukaca	salit apa lurnd berdukaca
3	innallahwainnallahroju ya allah	innallahwainnallahroju ya allah
4	tadi baca koment katanya lagi diel yaallah	tadi baca koment katanya lagi diel yaallah
5	please gals jangan mengging opini	please gals jangan mengging opini
6	fy high laura lula	fy high laura lula
7	yg pin respon di lid	yg pin respon di lid
8	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga

Gambar 6 Case Folding

Gambar 6 memperlihatkan hasil implementasi case folding, yakni proses untuk mengubah huruf kapital menjadi huruf kecil yang diterapkan Bagi pengkajian kebahasaan selanjutnya

4.2.4 Normalisasi Kata

Normalisasi kata merupakan salah satu tahapan penting dalam proses data cleaning pada text preprocessing yang bertujuan untuk menyeragamkan representasi kata dalam suatu korpus teks. Proses ini dilakukan dengan mengubah kata-kata tidak baku, singkatan, maupun variasi penulisan ke dalam bentuk baku sesuai dengan kaidah bahasa yang berlaku.

index	casefold_text	normal_text
0	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: 'Kataanya ngabalin? lu apaan makudnya?'	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: Kataanya ngabalin? lu apaan makudnya?
1	haahh	haahh
2	salit apa lurnd berdukaca	salit apa lurnd berdukaca
3	innallahwainnallahroju ya allah	innallahwainnallahroju ya allah
4	tadi baca koment katanya lagi diel yaallah	tadi baca koment katanya lagi diel yaallah
5	please gals jangan mengging opini	please gals jangan mengging opini
6	fy high laura lula	fy high laura lula
7	yg pin respon di lid	yg pin respon di lid
8	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga

Gambar 7 menampilkan contoh perihal

cara istilah nonformal dikonversi menjadi wujud baku selaras dengan acuan KBBI

4.2.5 Tokenizing

Tokenizing (tokenisasi) ialah satu di antara jenjang dalam *text preprocessing* yang berorientasi buat memfragmentasi naskah menjadi satuan-satuan yang kian mungil, yang dinamakan *token*. *Token* itu sanggup berwujud istilah, rangkaian kata, atau lambang spesifik yang akan digunakan sebagai dasar dalam proses analisis teks selanjutnya.

index	normal_text	token_text
0	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: 'Kataanya ngabalin? lu apaan makudnya?'	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: Kataanya ngabalin? lu apaan makudnya?
1	haahh	haahh
2	salit apa lurnd berdukaca	salit apa lurnd berdukaca
3	innallahwainnallahroju ya allah	innallahwainnallahroju ya allah
4	tadi baca koment katanya lagi diel yaallah	tadi baca koment katanya lagi diel yaallah
5	please gals jangan mengging opini	please gals jangan mengging opini
6	fy high laura lula	fy high laura lula
7	yg pin respon di lid	yg pin respon di lid
8	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga

Gambar 8 Tokenizing

Gambar 8 menunjukkan bagaimana sebuah aliran teks dipilah dan diurai menjadi komponen kecil, berawal dari diksi sampai tanda, serupa mozaik yang disusun dari fragmen-fragmen substansial..

4.2.6 Stopword Removal

Penghapusan stopwords ialah prosedur peniadaan istilah reguler yang tidak mempunyai esensi krusial dalam pengkajian narasi, seumpama konjungsi, preposisi, beserta pronomina. Tujuan dari tahapan ini adalah untuk mengurangi dimensi data serta menghilangkan kata-kata yang tidak memberikan kontribusi terhadap proses klasifikasi.

token_join	stopword_join	jumlah_sebelum	jumlah_sesudah
0	serius nanya ngabalin lu apaan? Tadi ada tat komen + n: 'Kataanya ngabalin? lu apaan makudnya?'	15	8
1	haahh	1	1
2	salit apa lurnd berdukaca	4	2
3	innallahwainnallahroju ya allah	3	3
4	tadi baca koment katanya lagi diel yaallah	7	4
5	please gals jangan mengging opini	5	4
6	fy high laura lula	4	4
7	yg pin respon di lid	5	2
8	jangan sepelekan penyakit gend guys jaga kesehatan dan raji olahraga	10	8
9		0	0
10		0	0
11	masih th pdli	3	2
12	baru banget nonton video tiktok lula sama reza...	13	11
13	ya udah	2	2
14	huh	1	1
15	mang siapa dia	3	1

Gambar 9 Stop Word Removal

Gambar 9 merupakan temuan pengeditan lewat penekanan efek penghilangan istilah-istilah yang tidak esensial.

4.2.7 Stemming

Stemming bersasaran guna menyatukan berbagai bentuk kata yang memiliki makna sama sehingga dapat meningkatkan akurasi dalam proses analisis teks.

Index	Sebelum Stemming	Setelah Stemming	Jumlah Kata Sebelum	Jumlah Kata Setelah
0	serius nanya ngelatin kat komen ex ngelatin mauludiyaa	serius nanya ngelatin kat komen ex ngelatin mauludiyaa	8	8
1	haahh	haahh	1	1
2	sih bawabaw	sih bawabaw	2	2
3	memblahwemalahingun ya alah	memblahwemalahingun ya alah	3	3
4	baca koment der suallah	baca koment der suallah	4	4
5	pkatan gas menging oati	pkatan gas menging oati	4	4
6	ty ngp hura hua	ty ngp hura hua	4	4
7	gan ngapi	gan ngapi	2	2
8	sepekan peripat panti panti jaga kesehatan ngp oimnaga	sepekan peripat panti panti jaga kesehatan ngp oimnaga	8	8
9	hi	hi	1	1
10	th pth	th pth	2	2
11	banget nonton video tiktok kula nica ang bkin beed at leat	banget nonton video tiktok kula nica ang bkin beed at leat	11	11
12	ya udah	ya udah	2	2
13	hai	hai	1	1
14	ming	ming	1	1
15	thut instagram kula	thut instagram kula	3	3
16	senas	senas	1	1
17	baca kn info neron dudaratan ya bnd thrah	baca kn info neron dudaratan ya bnd thrah	8	8
18	wah senas	wah senas	2	2
19	hah senas	hah senas	2	2
20	hemen hndap	hemen hndap	2	2

Gambar 10 Stemming

Gambar 10 merupakan hasil dari cara berbagai istilah dengan basis lema yang identik sanggup diidentifikasi

4.3 Pelabelan Data

Pelabelan data memiliki peran penting karena kualitas label yang diberikan akan sangat mempengaruhi performa model klasifikasi yang dihasilkan. Pelabelan dapat dilakukan secara manual oleh manusia (human annotation) atau secara otomatis menggunakan metode tertentu. Label yang diberikan biasanya adalah, sentimen positif, sentimen negatif, dan sentimen netral

```
df_balanced[['final_text', 'label']].head()
```

	final_text	label
4752	tenang lul	positif
9746	sedih banget ya allah	negatif
7560	cwokku jg ngerasain senang	positif
9880	lula aura baguss si arab	positif
8411	sakit banget nangisin bang arap wooyy	netral

Gambar 11 Pelabelan Data

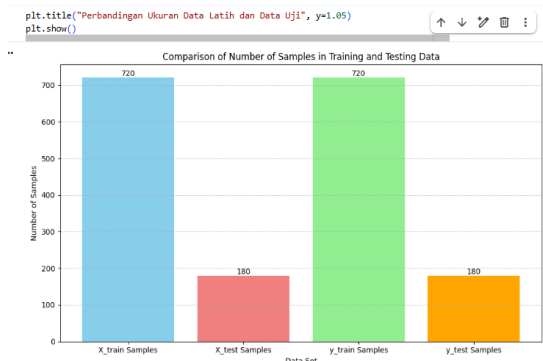
Gambar 11 menyajikan temuan pemberian identitas informasi, yang lantas digambarkan pada gambar diatas.

4.4 Implementasi Support Vector Machine

(isi penjelasan seperti dibawah tapi ruba ke Support vector machine - Guna mengelompokkan sentimen, Naive Bayes mengoperasikan perspektif bersandarkan probabilitas. Teknik ini berupaya mengestimasi intensitas kehadiran diksi khusus di tiap klasifikasi sentimen: positif, netral, dan negatif. Estimasi tersebut, yang berwujud probabilitas, lantas dipakai buat memastikan apakah suatu komentar berorientasi positif, netral, atau negatif. Kumpulan data pada riset tersebut diraih dari komentar pemakai Tiktok perihal perkara TPPO di Kamboja

yang sudah melewati fase penjernihan serta alih bentuk informasi. Di dalam kajian ini, dataset dipilah menjadi dua fragmen, yakni 80% un16/****uk materi edukasi serta 20% buat materi pengujian.

Gambar 12 memaparkan temuan proyeksi materi pengujian memakai teknik Naive Bayes Classifier bagi naskah komentar-nya)



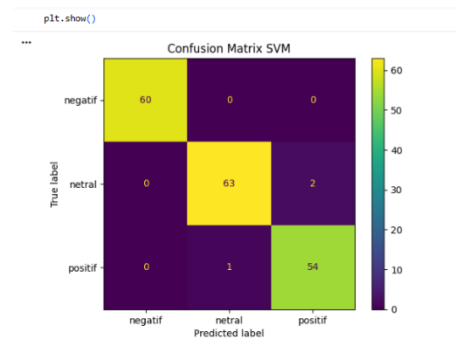
Gambar 12 Perbandingan Data Uji dan Data Latih

Gambar 12 menunjukkan komparasi kuantitas materi edukasi serta materi trial yang dioperasikan di observasi. Sejumlah 720 informasi dipakai selaku data latih serta 180 informasi sebagai data uji, baik untuk fitur (X) maupun label (y), dengan rasio pembagian 80:20.

4.5 Evaluasi Hasil

4.5.1 Confusion Matrix

Confusion matrix matrikmatrixmetrik penaksiran yang dimanfaatkan guna memperlihatkan totalitas estimasi yang akurat maupun keliru dari kerangka pemilahan tatkala mengolah materi yang sudah disematkan tanda.

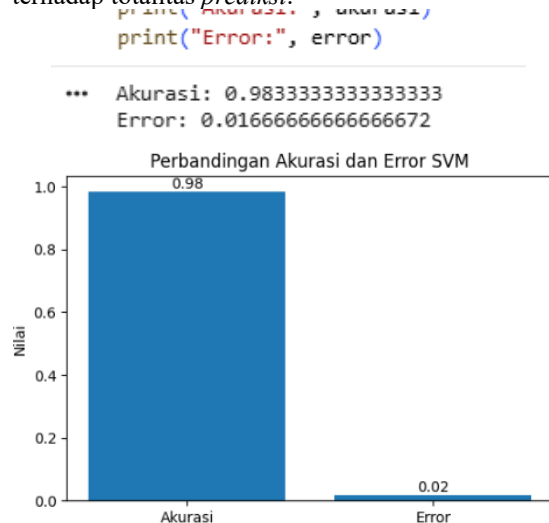


Gambar 3 Confusion Matrix

Gambar 13 memaparkan bahwasanya temuan *confusion matrix*-nya ialah *True* Positif 54, *True* Netral 63, serta *True* Negatif 60.

4.5.2 Accuracy dan Error

Akurasi ialah parameter intensitas *model* merumuskan *prediksi* yang akurat, lazimnya dikalkulasi lewat pembagian kuantitas *prediksi* yang valid terhadap keseluruhan *prediksi* yang disusun. *Error* merupakan antonim dari *akurasi*, yakni kuantitas *prediksi* yang keliru dari seluruh *prediksi* yang dieksekusi *model*. Umumnya diestimasi lewat pengomparasian jumlah *prediksi* yang luput terhadap totalitas *prediksi*.



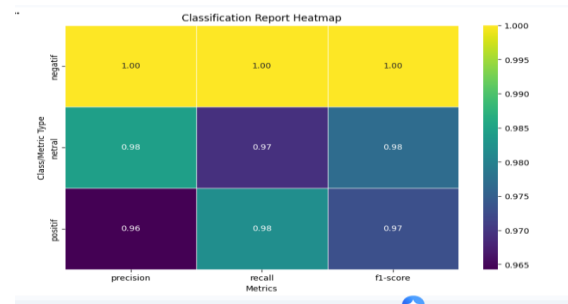
Gambar 14 Akurasi dan Error

Berdasarkan grafik “Perbandingan Akurasi dan Error SVM”, terlihat bahwa:

- Nilai akurasi model sangat tinggi, yaitu sekitar 0,98 (98%)
- Nilai error sangat rendah, yaitu sekitar 0,02 (2%).

4.5.3 Classification Report

Classification Report ialah *report* yang menjadi instrumen pemapar kualitas suatu *model* dalam memilah informasi. Di lingkupnya, mencakup parameter semacam *precision*, *recall*, *F1-score*, serta *support* atau kuantitas dokumentasi di tiap sentimen positif, netral, dan negatif. Dan lewat temuan *classification report* tersebut, kita dapat meninjau kinerja *model* Naive Bayes secara terperinci serta menetapkan defisit maupun keunggulan dalam pengategorian.

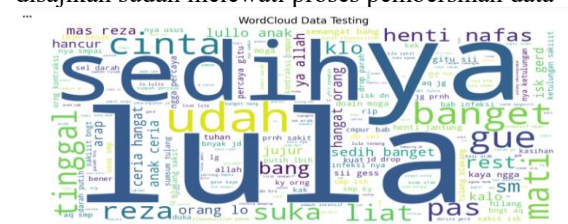


Gambar 15 Classification Report

Berdasarkan heatmap classification report, model SVM memaparkan hasil kerja yang amat prima di tiga kategori, yakni negatif, netral, dan positif, lewat angka *precision*, *recall*, serta *f1-score* yang semuanya berada di kisaran 0,96 – 1,00.

4.5.4 Wordcloud

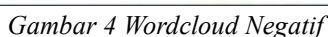
Wordcloud ialah sebuah metode penggambaran materi naskah yang dipakai guna melambangkan intensitas kehadiran istilah dalam suatu himpunan berkas atau korpus tulisan. Di dalam *wordcloud*, diksi yang hadir kian kerap bakal dipaparkan lewat dimensi *font* yang lebih masif, sementara istilah yang langka hadir diperlihatkan lewat ukuran kian mungil. Pemaparan grafis ini berorientasi guna menyuguhkan deskripsi permulaan bertalian kosakata dominan atau topik utama dalam data sehingga memudahkan proses eksplorasi data teks. Wordcloud umumnya dibangun berdasarkan hasil preprocessing seperti *tokenization*, *stopword removal*, serta *stemming*, hingga diksi yang disajikan sudah melewati proses pembersihan data



Gambar 16 Wordcloud Keseluruhan



Gambar 17 Wordcloud Positif



5.1 Kesimpulan

1. Proses analisis sentimen terhadap komentar

-
-
-
-
5. Temuan visualisasi wordcloud menunjukkan dominasi kata-kata yang berkaitan dengan ungkapan duka dan empati, sehingga dapat disimpulkan bahwa sebagian besar komentar pengguna TikTok cenderung mengandung sentimen emosional terhadap peristiwa yang dianalisis.
6. Secara keseluruhan, algoritma Support Vector Machine (SVM) terbukti efektif serta mampu memberikan hasil klasifikasi sentimen yang akurat mampu dioperasikan selaku metodologi guna menelaah aspirasi masyarakat pada sarana digital.

Bersandarkan pada temuan pengamatan yang sudah dilaksanakan, terdapat sejumlah saran yang bisa diajukan buat kemajuan riset mendatang:

1. Disarankan untuk menggunakan metode pelabelan data secara manual atau kombinasi antara manual dan otomatis guna meningkatkan kualitas label, karena pendekatan lexicon memiliki keterbatasan dalam memahami konteks bahasa.
2. Penelitian selanjutnya dapat membandingkan performa algoritma SVM dengan metode lain seperti Naive Bayes, Random Forest, maupun pendekatan deep learning guna memperoleh hasil yang lebih optimal.
3. Perlu dilakukan pengembangan pada tahap pre-processing, seperti penanganan emoji, bahasa tidak baku (slang), serta konteks kalimat, agar hasil analisis menjadi lebih akurat dan kontekstual.
4. Disarankan untuk menerapkan teknik evaluasi tambahan seperti cross-validation guna memastikan model tidak mengalami overfitting dan memiliki kemampuan generalisasi yang baik.
5. Penelitian selanjutnya diharapkan dapat menggunakan data yang lebih beragam dan berasal dari berbagai sumber agar hasil analisis lagi representatif terhadap opini publik secara luas.

Daftar Rujukan

- [1] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.
- [2] Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- [3] Fikri, A. R., & Setiawan, M. A. (2023). Analisis Sentimen Opini Publik Terhadap Program Bantuan Subsidi Upah (BSU) Di TikTok Dengan Pendekatan SVM. *Jurnal SITECH*, 6(2), 101–110.
- [4] Fitriani, E., & Nugroho, R. (2023). Analisis Sentimen Media Sosial TikTok Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(3), 211–220.
- [5] Goldberg, Y., & Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. *ArXiv Preprint*, 1–5.
- [6] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.
- [7] Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *European Conference on Machine Learning*, 137–142.
- [8] Kurniawan, M., & Setiawan, D. (2022). Analisis Sentimen Komentar YouTube Menggunakan Metode Support Vector Machine. *Jurnal RESTI*, 6(1), 55–63.
- [9] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- [10] Nurhayati, S. (2021). Implementasi Text Mining pada Analisis Sentimen Pengguna Media Sosial. *Jurnal Informatika*, 8(3), 201–210.
- [11] Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the International Conference on Language Resources and Evaluation*, 1320–1326.
- [12] Prameswari, D., & Kurniawan, I. (2022). Analisis Sentimen Terhadap Komentar Pada Video Viral TikTok Menggunakan Metode Naïve Bayes. *Jurnal Informatika Dan Komputer*, 5(2), 88–97.
- [13] Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python*. Packt Publishing.
- [14] Saputra, R., & Pratama, D. (2023). Analisis Sentimen Aplikasi TikTok Menggunakan Algoritma Naïve Bayes dan Support Vector Machine. *Jurnal Tekno Kompak*, 17(1), 45–54.
- [15] Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1–47.
- [16] Suartika, I. W., & Arsi, P. (2022). Klasifikasi Sentimen Ulasan Aplikasi Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Teknologi Informasi Asia*, 16(1), 77–86.
- [17] Syahputra, A., Rizki, M., & Hidayat, F. (2023). Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE. *Jurnal INSTEK*, 8(1), 33–42.
- [18] Wibowo, A., & Firmansyah, R. (2022). Penerapan TF-IDF dan Support Vector Machine untuk Analisis Sentimen Media Sosial Twitter. *Jurnal Teknologi Informasi Dan Komunikasi*, 9(2), 120–129.