

BAB 2

TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah sebuah bidang studi yang menganalisis pendapat, sentimen, penilaian dan emosi seseorang terhadap suatu barang, orang maupun sebuah peristiwa. Ada beberapa penamaan terhadap studi ini, yaitu analisis sentimen, penambangan opini, ekstraksi opini dan penambangan sentimen dimana semuanya sekarang berada dibawah ranah analisis sentimen. Analisis sentimen bertujuan untuk menentukan suatu polaritas dalam sebuah review terhadap orang maupun peristiwa dalam kelas positif atau negatif dari pendapat yang berbentuk dokumen atau tulisan[5].

2.2. Satire

Satire adalah gaya bahasa yang dipakai dalam kesusastraan untuk menyatakan sindiran terhadap suatu keadaan atau seseorang. Menurut Keraf seperti dikutip dalam Tarigan, satire adalah ungkapan yang menertawakan atau menolak sesuatu. Bentuk ini tidak harus bersifat ironis. Satire mengandung kritik tentang kelemahan manusia. Tujuan utamanya adalah agar diadakan perbaikan secara etis maupun estetis[6]. Satire merupakan karikatur, parodi, sindiran[7].

Satire biasanya disampaikan dalam bentuk ironi, sarkasme dan parodi. Satire adalah gaya bahasa yang dipakai dalam kesusastraan untuk menyatakan sindiran, hinaan terhadap suatu keadaan atau seseorang. Sedangkan arti kata hinaan adalah merendahkan, memandang rendah, memburukkan nama baik orang, menyinggung perasaan (seperti memaki-maki, menistakan)[2].

Kamu kebanyakan makan cabai, ya? Mulutmu pedas sekali.

Gambar 2.1 Contoh Kalimat Satire

Gunakan kepalamu untuk berpikir, bukan mulut untuk komentar tidak jelas.

Gambar 2.2 Contoh Kalimat Sarkas

2.3. Preprocessing

Tahap *preprocessing* dilakukan sebelum dilakukannya proses klasifikasi, tujuan dari tahap *Preprocessing* ini dilakukan untuk membersihkan data dari kata, simbol dan huruf yang tidak diperlukan agar menyeragamkan bentuk kata sehingga dapat mengurangi volume kata. Karena sebelum dilakukan tahap *Preprocessing* dataset memungkinkan memiliki noise yang tinggi yang dapat mempengaruhi proses klasifikasi[8].

2.3.1 Case Folding

Case folding adalah proses penyeragaman bentuk huruf atau mengubah ukuran semua huruf menjadi sama besar. Dalam proses ini semua kata dari komentar yang terdapat huruf besar akan diubah menjadi huruf kecil (*lower case*). Proses ini bertujuan untuk memudahkan pencarian, karena tidak semua kata konsisten dalam penggunaan huruf kapital[9].

2.3.2 Cleaning

Cleaning adalah proses penghilangan karakter symbol yang dianggap sebagai noise. Karakter yang dimaksud seperti koma(,) titik(.) kutip(') kutip dua(") tanda tanya(?) tanda seru(!) dan sebagainya. Proses *cleaning* dilakukan untuk membersihkan karakter simbol yang tidak memiliki arti. Selain karakter simbol, angka, hashtag(#), at(@), URL dalam komentar pun dihilangkan karena tidak berhubungan dengan proses analisis sentimen[10].

2.3.3 Tokenizing

Tokenizing adalah proses pemisahan atau pemotongan semua kata yang ada pada suatu kalimat atau dokumen dan mengubahnya menjadi kumpulan token atau term. Token yang dihasilkan akan digunakan sebagai pembanding pada proses klasifikasi. *Tokenizing* merupakan proses memisahkan input text menjadi potongan yang memiliki arti[11].

2.3.4 Normalization

Normalization merupakan proses perbaikan kata-kata yang salah eja atau disingkat dalam bentuk tertentu[12]. Dalam sebuah komentar pasti terdapat kata-kata yang disingkat ataupun typo seperti “tdk” yang seharusnya “tidak” ataupun “iy” yang seharusnya “iya”. Proses ini dilakukan dengan menyusun kamus kata atau memanipulasi singkatan tersebut. Sederhananya, proses ini dilakukan untuk memperbaiki kata yang salah ketik dalam penulisan.

2.3.5 Convert Negation

Convert Negation adalah proses mempertegas kata yang bermakna negasi. Dalam sebuah komentar sangat memungkinkan terdapat negasi yang dapat membalikan arti dari suatu kata yang akan mengubah sentimen yang tadinya positif menjadi negatif maupun sebaliknya[13]. Jika pada komentar mengandung kata negasi, maka akan digabungkan dengan kata setelahnya.

2.3.6 Stopword Removal

Stopwords removal merupakan tahapan untuk membuang kata-kata yang tidak dibutuhkan[12]. *stopword removal* adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna, sehingga tidak akan mempengaruhi sentimen. *Stop words* biasanya seperti kata kerja, kata sifat, kata keterangan, dan kata sambung.

2.4. Unigram

Metode N-Gram adalah sekumpulan n-kata yang muncul dalam urutan pada kalimat atau kumpulan teks. Pemodelan N-Gram adalah pendekatan identifikasi dan analisis fitur yang populer digunakan dalam pemodelan bahasa dan bidang pemrosesan bahasa alami. *Characters* dan *Word* merupakan model N-Gram yang paling banyak digunakan dalam kategori teks. N-Gram dalam analisis sentimen membantu menganalisis sentimen teks atau dokumen. Karakteristik pada N-Gram adalah sebagai berikut:

1. Walaupun terdapat kesalahan tekstual, N-Gram masih dapat berfungsi dengan baik.

2. Membutuhkan penyimpanan yang sederhana dan berjalan dengan efisien.
3. Waktu proses yang dibutuhkan relatif cepat.

Dalam bahasa Indonesia terdapat banyak frase yang tidak hanya terdiri dari satu kata oleh karena itu N-Gram sangat dibutuhkan. Unigram merupakan bagian dari n-gram, yaitu potongan kata sebanyak n berdasarkan pada urutan sekuensial teks string[15]. Unigram merupakan n- gram dengan nilai n adalah satu n-gram berukuran satu, Contoh: terdapat sebuah teks “Universitas Komputer Indonesia”, maka unigram dari teks tersebut adalah “Universitas”, “Komputer”, “Indonesia”.

2.5. Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode dalam pembobotan kata dengan mengintegrasikan *tf* (*term frequency*) dan *idf* (*inverse document frequency*). Metode ini digunakan untuk memberikan nilai terhadap suatu kata berdasarkan tingkat kepentingan terhadap dokumen atau kategori dalam suatu kumpulan dokumen. TF-IDF terdapat dua bagian yaitu TF dan IDF, dimana perhitungan TF (*Term Frequency*) menunjukkan jumlah munculnya suatu kata pada suatu dokumen/teks, sedangkan DF (*Document Frequency*) merupakan jumlah dari dokumen/teks yang terdapat suatu kata/term[12]. Nilai Tf-Idf yang diperoleh dengan mengalikan nilai Tf dengan nilai Idf[14]. Persamaan TF-IDF dapat dirumuskan dengan persamaan 2.1 berikut.

$$W(i,j) = tf_{(i,j)} * \log\left(\frac{D}{DF_j}\right) \quad (2.1)$$

Dimana:

- $W(i,j)$: bobot kata j terhadap dokumen i
 $tf_{(i,j)}$: banyaknya kemunculan kata j pada dokumen i
 D : Jumlah Dokumen
 DF_j : jumlah dokumen yang berisi kata j.

2.6. *Support Vector Machine*

Banyak penelitian yang menggunakan berbagai macam algoritma dalam proses klasifikasi, SVM merupakan metode klasifikasi yang membagi menjadi dua kelas yaitu positif dan negatif. SVM memiliki sebuah keunggulan pada efektivitas saat melakukan klasifikasi yang dapat menangani dokumen dengan input ruang dimensi yang tinggi dan dapat menghilangkan fitur-fitur yang tidak relevan. Tetapi SVM memiliki salah satu kelemahan yaitu algoritma pelatihan dan pengkategorian relatif kompleks, membutuhkan waktu yang cukup tinggi, dan banyak menggunakan memori pada proses pelatihan dan klasifikasi[16].

Support Vector Machine adalah teknik yang relatif baru untuk membuat prediksi baik dalam klasifikasi maupun regresi dan SVM adalah solusi global yang optimal dan menghindari *dimensionality*. SVM berada di kelas yang sama dengan ANN dalam hal fitur dan status masalah yang dapat dipecahkan. Keduanya termasuk dalam kategori *supervised learning*, di mana dalam implementasinya memerlukan tahap *training* dan dilanjutkan dengan tahap *testing*. Teknik SVM ini berusaha menemukan fungsi klasifikasi terbaik yang dapat memisahkan dua dataset dari dua kelas yang berbeda. Dalam beberapa penelitian ditunjukkan bahwa SVM adalah metode yang efisien [17][18][19].

Support Vector Machine adalah metode klasifikasi untuk mencari *Maximum Marginal Hyperplane* (MMH) atau pemisah terbaik untuk menciptakan pemisahan maksimal untuk semua kelas [20]. Margin bisa didefinisikan sebagai jarak terpendek dari sebuah *hyperplane* terhadap satu sisi dari margin itu sama dengan jarak *hyperplane* dengan sisi margin lainnya, dengan catatan kedua margin tersebut dalam posisi paralel dengan *hyperplane* [29]. Berikut merupakan konsep SVM yang dimulai dengan persamaan garis lurus. Dengan rumus persamaan (2.2) berikut:

$$y = ax + b \quad (2.2)$$

Persamaan garis lurus tersebut dituliskan pada persamaan (2.11) berikut:

$$y - ax - b = 0 \quad (2.3)$$

Dimana:

a = kemiringan atau gradien (m)

b = bilangan konstanta, a dan b merupakan bilangan real dan a tidak nol.

Dalam SVM, secara general sebuah *hyperplane* dinyatakan pada persamaan (2.4) berikut:

$$w \cdot x + b = 0 \quad (2.4)$$

Keterangan:

w = nilai dari bidang normal

x = data input

b = posisi bidang relatif terhadap pusat koordinat, dimana skalar b bisa bernilai negatif, nol, maupun positif.

Pencarian *hyperplane optimum* yang memaksimalkan margin dapat dipandang sebagai sebuah masalah *Quadratic Programming* (QP), yaitu mencari titik minimal dari persamaan (2.5) dengan memperhatikan batasan pada persamaan (2.6) berikut:

$$\min \tau(w) = \frac{1}{2} \|w\|^2 \quad (2.5)$$

$$y_i(x_i \cdot w + b) - 1 \geq 0, \forall i \quad (2.6)$$

Masalah Lagragian dijelaskan pada persamaan (2.7) untuk pengklasifikasian

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i(x_i \cdot w + b) - 1), i = 1, 2, \dots, l \quad (2.7)$$

Kemudian untuk memaksimalkan L terhadap α_i dijelaskan pada persamaan (2.8) dengan memperhatikan batasan pada persamaan (2.9).

$$\max_{\alpha} L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2.8)$$

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \sum_{i,i=1}^l \alpha_i y_i = 0 \quad (2.9)$$

Setelah mendapatkan *Support Vector* dengan α_i bernilai positif, maka selanjutnya menghitung w yang dapat dilihat pada persamaan (2.10) dan menghitung b yang dapat dilihat pada persamaan (2.11).

$$w = \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (2.10)$$

$$b = y_k - w^T x_k \quad (2.11)$$

Berikutnya untuk menghitung kelas \bar{x} , dapat dilihat pada persamaan (2.12)

$$F(\bar{x}) = W^T x + b \quad (2.12)$$

Berikut ini adalah beberapa fungsi kernel yang umum digunakan, yaitu:

Tabel 2.1. Kernel Trik SVM

Kernel Type	Formula
Linier	$K(x, x_k) = x_k^T x$
Polynomial	$K(x, x_k) = (x_k^T x + 1)^d$
Radial Basis Function	$K(x, x_k) = \exp\{-\ x - x_k\ _2^2 / \sigma^2\}$
Sigmoid Kernel	$K(x, x_k) = \tanh [k x_k^T x + \theta]$

2.7. Confusion Matrix

Evaluasi performansi dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kebenaran dari sistem. Metode yang digunakan untuk evaluasi adalah *confusion matrix*. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Dalam pengujian keakuratan hasil pencarian akan dievaluasi nilai *recall*, *precision*, dan akurasi. Dimana *precision* mengevaluasi kemampuan sistem untuk menemukan peringkat yang paling relevan, dan didefinisikan sebagai presentase dokumen yang

di-*retrieve* dan benar-benar relevan terhadap *query*. *Recall* mengevaluasi kemampuan sistem untuk menemukan semua item yang relevan dari koleksi dokumen dan didefinisikan sebagai presentase dokumen yang relevan terhadap *query*. Sedangkan akurasi sendiri merupakan perbandingan kasus yang diidentifikasi benar dengan jumlah seluruh kasus[23].

Confusion Matrix menganalisis seberapa baik klasifikasi yang telah dilakukan terhadap kelas aktual dan kelas hasil prediksi *matrix*. Hasil dari *confusion matrix* berupa nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Evaluasi yang akan digunakan yaitu *accuracy* [21].

Tabel 2.2 Tabel *Confusion Matrix*

<i>Confusion Matrix</i>	Kelas Prediksi			
		Positif	Satire	Negatif
Kelas Aktual	Positif	TP	FPS	FPN
	Satire	SFP	TS	SFN
	Negatif	NFP	FNS	TN

Nilai *true positive* (TP) dan *true negative* (TN) adalah hasil klasifikasi yang benar. Nilai *false positive* (FP) adalah data yang bernilai negatif namun terderekasi sebagai data positif sedangkan *false negative* (FN) adalah data yang bernilai positif namun terdeteksi sebagai data negatif. *Confusion Matrix* menunjukkan tingkat akurasi dari proses klasifikasi yang telah dilakukan. Tingkat akurasi menunjukkan proporsi jumlah prediksi benar.

$$Akurasi = \frac{TP+TS+TN}{Total} \quad (2.13)$$

2.8. Penelitian-penelitian Terkait

Tabel 2.3 Penelitian-penelitian Terkait

Review Literatur Pertama[1]	
Judul Paper	Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter
Penulis	Ghulam Asrofi Buntoro
Judul Jurnal/Proceeding	Integer Journal
Tahun Penerbitan	2017
Masalah utama yang diangkat	Pemilihan Gubernur DKI Jakarta 2017 ramai diperbincangkan khususnya di media sosial Twitter. Semua orang bebas berpendapat atau beropini tentang calon Gubernur DKI Jakarta 2017 sehingga memunculkan banyak opini, tidak hanya opini yang positif atau netral tapi juga yang negatif. Media sosial khususnya Twitter sekarang ini menjadi salah satu tempat promosi atau kampanye yang efektif dan efisien. Penelitian ini diharapkan dapat bermanfaat membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif, netral atau negatif.
Metode Ekstraksi	<i>Lexicon Based</i>
Metode Klasifikasi	<i>Naïve Bayes Classifier (NBC)</i> dan <i>Support Vector Machine (SVM)</i>

<p>Hasil Penelitian dan Kesimpulan</p>	<p>Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen Twitter terhadap calon Gubernur DKI Jakarta 2017. Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada twit opini Bahasa Indonesia yang ada di Twitter. Nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi <i>Naïve Bayes Classifier</i> (NBC) untuk klasifikasi data AHY, dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6%. Dalam penelitian ini juga dapat diketahui metode klasifikasi <i>Naïve Bayes Classifier</i> (NBC) lebih tinggi akurasinya untuk klasifikasi sentimen Tweet Bahasa Indonesia dibandingkan dengan metode klasifikasi <i>Support Vector Machine</i> (SVM).</p>
<p>Masalah yang ditemukan</p>	<p>-</p>

Review Literatur Kedua[22]	
<p>Judul Paper</p>	<p>Implementasi Text Mining Untuk Analisis Opini Masyarakat Terhadap Kinerja Layanan Transportasi Online Dengan Analisis Faktor</p>
<p>Penulis</p>	<p>Immanuel Olive Djaja Putra, Kestriia Rega Prilianti, Paulus Lucky Tirma Irawan</p>
<p>Judul Jurnal/Proceeding</p>	<p>Jurnal SimanteC, Volume 8, No 2</p>

Tahun Penerbitan	2020
Masalah utama yang diangkat	Analisis sentimen masyarakat terhadap Gojek dan Grab secara manual membutuhkan banyak waktu dan tenaga yang dikeluarkan karena informasi yang terdapat media sosial bergerak dengan cepat, maka diperlukan teknik analisis sentimen secara otomatis agar dapat mempermudah penyedia jasa untuk memperbaiki maupun mengembangkan kualitas layanannya.
Metode Ekstraksi	Principal Component Analysis (PCA), TF-IDF
Metode Klasifikasi	Algoritma <i>Naïve Bayes Classifier</i>
Hasil Penelitian dan Kesimpulan	<p>Hasil : Pada proses perhitungan akurasi objek Grab didapatkan bahwa sebanyak 113 tweet dari 152 tweet berhasil diklasifikasikan secara benar. Pada objek Gojek sebanyak 389 tweet dari 565 tweet berhasil diklasifikasikan secara benar, dengan begitu nilai akurasi pada data uji grab adalah 74.34%, pada data uji gojek adalah 68.84% dan total akurasi rata-rata dari kedua objek tersebut adalah 71.59%.</p> <p>Kesimpulan : Seleksi fitur yang digunakan untuk menghapus jumlah kata yang kurang dari tiga berjalan dengan baik dan Metode PCA dapat mengelompokkan kata ke dalam beberapa faktor.</p>
Masalah yang ditemukan	Perlu menambahkan data latih untuk menyeimbangkan data antara sentimen positif dan sentimen negatif agar dapat meningkatkan hasil akurasi pada proses klasifikasi sentimen.

Review Literatur Ketiga[23]	
Judul Paper	<i>Sentiment Analysis of Product Reviews using Support Vector Machine Learning Algorithm</i>
Penulis	Esha Tyagi dan Arvind Kumar Sharma
Judul Jurnal/Proceeding	<i>Indian Journal of Science and Technology</i> , Vol 10(35)
Tahun Penerbitan	2017
Masalah utama yang diangkat	Berbagai ulasan <i>pengguna</i> untuk produk, perusahaan, merek, individu, forum, dan film tertentu, dll. telah banyak membantu dalam menilai persepsi orang.
Metode Ekstraksi	TF-IDF
Metode Klasifikasi	<i>Support Vector Machine (SVM)</i>
Hasil Penelitian dan Kesimpulan	<i>Support Vector Machine (SVM)</i> sebagai metode klasifikasinya seperti “ <i>Sentiment Analysis of Product Reviews using Support Vector Machine Learning Algorithm</i> ” mengatakan Algoritma klasifikasi <i>Support Vector Machine (SVM)</i> 89,98% lebih akurat dibandingkan algoritma lainnya. Keakuratan yang diperoleh akan lebih ditingkatkan dengan memasukkan lebih banyak bentuk kalimat. Akhirnya, disimpulkan bahwa algoritma SVM lebih baik.
Masalah yang ditemukan	-

Review Literatur Keempat[24]	
Judul Paper	Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter
Penulis	Fajar Fathur Rachman, Setia Pramana
Judul Jurnal/Proceeding	Indonesian of Health Information Management Journal
Tahun Penerbitan	2020
Masalah utama yang diangkat	Untuk melakukan percepatan penanganan penyebaran COVID-19 di Indonesia, Pemerintah Republik Indonesia telah mengeluarkan wacana vaksinasi untuk masyarakat Indonesia pada akhir tahun 2020 mendatang. Wacana tersebut dinilai kontroversial sehingga mengundang banyak kalangan untuk memberikan pendapatnya di berbagai media. Pendapat tersebut haruslah dipertimbangkan sebagai bahan evaluasi sehingga rencana vaksinasi yang akan dilakukan dapat berjalan dengan baik. Dengan memanfaatkan data dari media sosial twitter, penelitian ini bertujuan untuk menganalisis respon masyarakat terhadap wacana vaksinasi dengan cara mengklasifikasikan respon tersebut ke dalam respon positif dan negatif.
Metode Ekstraksi	Lexicon Based
Metode Klasifikasi	Latent Dirichlet Allocation (LDA)
Hasil Penelitian dan Kesimpulan	Masyarakat lebih banyak memberikan respon yang bersentimen positif terhadap vaksin COVID-19 dibandingkan dengan respon yang bersentimen negatif.

	Model LDA yang dibangun dapat menangkap berbagai macam topik pembicaraan masyarakat di media sosial twitter terkait vaksin COVID-19 seperti pembicaraan masyarakat mengenai vaksin merah putih, sertifikasi halal vaksin, uji layak pakai vaksin, harga vaksin, sampai pembicaraan umum masyarakat seperti fungsi & objek vaksinasi[24].
Masalah yang ditemukan	-

Review Literatur Kelima[25]	
Judul Paper	Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi
Penulis	Angelina Puput Giovani, Ardiansyah, Tuti Haryanti, Laela Kurniawati, Windu Gata
Judul Jurnal/Proceeding	Jurnal TEKNOINFO
Tahun Penerbitan	2020
Masalah utama yang diangkat	E-learning merupakan pembelajaran berbasis elektronik dengan menggunakan komputer atau berbasis komputer. Salah satu aplikasi e-learning yang banyak dikenal saat ini adalah aplikasi Ruang Guru. Salah satu cara untuk

	mengetahui keberhasilan suatu aplikasi adalah dengan melakukan analisis sentimen terhadap aplikasi tersebut.
Metode Ekstraksi	<i>Naive Bayes</i> (NB), <i>Support Vector Machine</i> (SVM), <i>K-Nearest Neighbour</i> (K-NN)
Metode Klasifikasi	<i>feature selection</i> PSO
Hasil Penelitian dan Kesimpulan	Penggunaan <i>feature selection</i> PSO dalam algoritma klasifikasi dapat meningkatkan performa dan akurasi. Hasil dari pengujian secara keseluruhan algoritma SVM (PSO) memiliki nilai akurasi dan performa yang paling tinggi jika dibandingkan dengan NB, SVM, KNN, NB (PSO), SVM (PSO), dan K-NN (PSO). Sedangkan algoritma NB, tanpa PSO maupun dengan PSO, mendapatkan nilai UAC dengan klasifikasi gagal[25].
Masalah yang ditemukan	-

Review Literatur Keenam[26]	
Judul Paper	Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode <i>Naive Bayes</i>
Penulis	Didik Garbian Nugroho, Yulison Herry Chrisnanto, Agung Wahana
Judul Jurnal/Proceeding	Prosiding SNST ke-7
Tahun Penerbitan	2016

Masalah utama yang diangkat	Analisis sentimen pada jasa ojek online merupakan proses mengekstraksi pendapat, sentimen, evaluasi, dan emosi orang tentang pelayanan ojek online yang tertulis. Di media sosial masyarakat mengeluarkan beragam opini tentang pelayanan dari transportasi ini dengan jumlah yang banyak, sehingga terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral.
Metode Ekstraksi	-
Metode Klasifikasi	<i>Naïve Bayes Classifier</i> (NBC)
Hasil Penelitian dan Kesimpulan	Penelitian ini menghasilkan sebuah sistem analisis sentimen pada jasa ojek online menggunakan metode <i>Naïve Bayes</i> . Proses penelitian ini menggunakan masukan berupa tweet mention jasa ojek online di Twitter. Hasil dari pengujian, sistem mampu mengklasifikasi sentimen menggunakan <i>Naïve Bayes</i> dengan akurasi yang dihasilkan sebesar 80% berdasarkan 800 data tweet yang terdiri atas 300 data latih dan 500 data uji[26].
Masalah yang ditemukan	-

Review Literatur Ketujuh[27]	
Judul Paper	Literatur Review: Metode Klasifikasi Pada Sentimen Analisis
Penulis	Mona Cindo, Dian Palupi Rini, Ermatita

Judul Jurnal/Proceeding	Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)
Tahun Penerbitan	2019
Masalah utama yang diangkat	Kajian ini membahas tentang metode dan fitur ekstraksi yang memungkinkan untuk dapat melakukan sentimen analisis.
Metode Ekstraksi	<i>Lexicon Based</i>
Metode Klasifikasi	<i>Naïve Bayes, Support Vector Machine, Logistic Regression</i>
Hasil Penelitian dan Kesimpulan	Pada beberapa penelitian sebelumnya lebih banyak menggunakan Twitter untuk mengumpulkan dataset. Metode klasifikasi yang paling sering digunakan adalah <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> . Namun, hasil yang terbaik ada pada metode logistic regression dan lexical-based. Hasil tersebut mungkin berdampak dari pemilihan fitur ekstraksi yang tepat dan penyusunan dataset yang baik[27].
Masalah yang ditemukan	-