

Jurnal_Ricky_Renaldo_A_161080 200182.docx

by

Submission date: 27-Mar-2023 09:11AM (UTC+0700)

Submission ID: 2047422081

File name: Jurnal_Ricky_Renaldo_A_161080200182.docx (707.4K)

Word count: 4128

Character count: 23977

Analysis Of Sentiment Comments On JNE Using K-Nearest Neighbor (KNN) Method On Twitter Social Media

Analisis Sentimen Komentar Terhadap JNE Menggunakan Metode K-Nearest Neighbor (KNN) Pada Media Sosial Twitter

Ricky Renaldo Arisandi¹, Ir. Sumarno., MM.², Hamzah Setiawan, S.Kom., M.Kom.³

^{1,2,3}Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

Email : cassaduabelas@gmail.com¹, sumarno@umsida.ac.id², hamzah@umsida.ac.id³

Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo

Abstract. Social media has now become a public space and virtual criticism along with the widespread use of smartphones, especially Twitter. Netizens use it as an effective communication channel because it is not that difficult to use. By one click, millions of connected netizens can see it easily without barriers and boundaries. The results of the public opinion can be used as data for sentiment analysis in the form of comments and used as material for text mining analysis. With the commentary text data, there are various kinds of opinions given by the public through social media, for example, on Twitter. This study aims to accelerate efforts to obtain more accurate information about the mood of the mass media, and the results of this information are more accurate if the services provided are negative, positive and neutral. Based on the background above, the authors propose that the research to be discussed is about Comment Sentiment Analysts Against JNE Using the K-Nearest Neighbor (KNN) Method on Twitter.

Keywords - Social media; Twitter; Sentiment Analysis; Mining Text; K-Nearest Neighbor; KNN.

Abstrak. Saat ini media sosial telah menjadi ruang publik tanpa batas ruang dan waktu yang ditandai dengan banyaknya penggunaan smartphone terutama Twitter. Netizen menggunakan sebagai sarana komunikasi yang efektif karena sangat mudah penggunaannya. Hasil opini masyarakat tersebut dapat digunakan sebagai data untuk analisa sentimen yang berupa komentar dan dijadikan bahan untuk analisa teks mining. Dengan data teks komentar tersebut terdapat berbagai macam opini yang diberikan oleh masyarakat melalui media sosial contohnya seperti di media sosial twitter. Penelitian ini bertujuan untuk mempermudah dalam memperoleh informasi yang lebih akurat tentang mood media massa, dan hasil informasi tersebut lebih akurat jika layanan yang diberikan negatif, positif dan netral. Berdasarkan latar belakang diatas, penulis mengusulkan penelitian yang akan dibahas yaitu tentang Analis Sentimen Komentar Terhadap JNE Menggunakan Metode K-Nearest Neighbor (KNN) pada Media Sosial Twitter.

Kata Kunci – Media Sosial; Twitter; Analisa Sentimen; Teks Mining; K-Nearest Neighbor; KNN.

I. PENDAHULUAN

PT. Tiki Jalur Nugraha Eka Kurir atau biasa disebut JNE Express merupakan salah satu perusahaan penyedia jasa pengiriman barang terbesar dengan jaringan terluas di Indonesia. Lingkup bisnis yang di jalani oleh PT. JNE EXPRESS adalah pengiriman (*courier*), penyimpanan dan pendistribusian (*logistic and distribution*), transportasi (*moving, trucking, air sea cargo*), kepabean (*custom clereance*), penjemputan bandara (*escort service*), pengiriman uang (*money transfer*) [1]. Semakin banyaknya peminat pengguna layanan tersebut semakin banyak pula opini yang didapatkan dari masyarakat, baik itu tentang fasilitas ataupun lainnya.

Media sosial telah menjadi ruang publik tanpa batas ruang dan waktu yang ditandai dengan banyaknya penggunaan *smartphone* terutama Twitter. Netizen menggunakan sebagai sarana komunikasi yang efektif karena sangat mudah dalam penggunaannya. Sekali klik, jutaan netizen yang terhubung dapat melihat dengan mudah tanpa batas [2]. Hasil opini masyarakat tersebut dapat digunakan sebagai data untuk analisa sentimen yang berupa komentar dan dijadikan bahan untuk analisa teks mining. Dengan data teks komentar tersebut terdapat berbagai macam opini yang diberikan oleh masyarakat melalui media sosial contohnya seperti di media sosial twitter. PT Tiki Jalur Nugraha Eka Kurir (JNE) memproyeksikan pertumbuhan pengiriman paket hingga akhir tahun 2022 dapat konsiten sekitar 30%-40%. Optimisme ini didukung oleh perkembangan *e-commerce* yang begitu signifikan, sehingga dapat turut meningkatkan permintaan pengiriman paket ke depan [3].

JNE mempunyai akun resmi di twitter yaitu bernama JNE Express, dalam akun tersebut terdapat berbagai *tweet* dan komentar masyarakat yang telah diberikan, sehingga perlu adanya teknologi yang tepat untuk melakukan analisa sentimen. Metode yang dapat digunakan untuk analisa sentimen adalah metode KNN. Keuntungan menggunakan metode ini yaitu tangguh terhadap training data yang noisy dan efektif apabila datanya besar serta metode yang mudah dipahami dan mudah diimplementasikan. Penelitian ini bertujuan untuk mempermudah dalam memperoleh informasi yang lebih akurat tentang mood media massa, dan hasil informasi tersebut lebih akurat jika layanan yang diberikan negatif, positif dan netral.

Berdasarkan latar belakang diatas, penulis mengusulkan penelitian yang akan dibahas yaitu tentang Analisis Sentimen Komentar Terhadap JNE Menggunakan Metode K-Nearest Neighbor (KNN) pada Media Sosial Twitter menggunakan 100 data.

A. JNE

PT. Tiki Jalur Nugraha Eka Kurir didirikan oleh H. Soeprato Suparno pada tanggal 26 November 1990. Perusahaan ini merupakan divisi dari PT Citra van Titipan Kilat (TIKI) yang bertugas untuk mengelola jaringan kurir internasional. Berawal dari delapan orang dan modal 100 juta rupiah, JNE memulai usahanya yang terpusat pada penanganan kegiatan kepabeanan, impor kiriman barang, dokumen serta pengantarannya dari luar negeri ke Indonesia [4].

B. Twitter

Twitter merupakan sebuah layanan jejaring sosial dan bisa disebut juga *microblog daring* yang memungkinkan penggunaanya untuk berkirim dan membaca pesan berbasis teks yang tidak lebih dari 280 karakter dalam sebuah tweet [5].

C. Analisa Sentimen

Analisa sentimen bagian dari opinion mining merupakan proses memahami, mengestrak dan mengolah data tekstual yang secara otomatis untuk memperoleh sebuah informasi. Analisa sentimen sering disebut juga opinion mining (penambangan opini) adalah koputasi untuk mengenali dan mengekspresikan opini, sentimen evaluasi, sikap, emosi, subjektifitas, penilaian serta pandangan dalam suatu teks [6].

D. Database

Database atau basis data adalah kumpulan dari beberapa informasi yang disimpan di dalam komputer secara sistematis sehingga dapat diperiksa menggunakan suatu program komputer untuk memperoleh informasi dari basis data tersebut [7].

E. Flowchart

Flowchart merupakan penggambaran secara grafik dari langkah-langkah dan urutan prosedur suatu program. Biasanya mempengaruhi penyelesaian masalah yang khususnya perlu dipelajari dan dievaluasi lebih lanjut [8].

F. Python

Python merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaan besar maupun para developer untuk mengembangkan berbagai macam aplikasi berbasis desktop, web dan mobile. Python diciptakan oleh Guido van Rossum di Belanda pada tahun 1990 dan namanya diambil dari acara televisi kesukaan Guido Monty Python's Flying Circus. Van Rossum mengembangkan Python sebagai hobi, kemudian Python menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak intuitif dan memiliki pustaka yang luas [9].

G. Text Mining

Text mining adalah sebuah proses yang digunakan untuk mengeksplorasi data berupa teks, dimana sumber data tersebut biasanya didapatkan dari dokumen [10]. Text Mining bertujuan untuk menemukan informasi berharga yang tersembunyi baik dari sumber informasi terstruktur dan tidak terstruktur. Text Mining merupakan suatu penemuan baru yang sebelumnya informasinya tidak diketahui. Informasi yang diekstrak dari berbagai sumber daya tertulis dilakukan secara otomatis. Elemen kuncinya adalah menghubungkan beberapa informasi yang diekstraksi menjadi satu sehingga dapat membentuk fakta baru atau hipotesis baru untuk dieksplorasi lebih lanjut [11]. Tugas text mining antara lain, pengkategorian teks (*text categorization*) dan pengelompokan teks (*text clustering*).

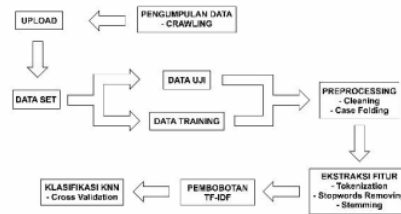
II. METODOLOGI PENELITIAN

A. Identifikasi Masalah

Pengklasifikasi K-Nearest Neighbor (KNN) dapat di klasifikasikan dengan mempelajari "kumpulan dari model contoh dokumen yang sebelumnya diklasifikasikan" Dengan pembobotan TF-IDF dalam mengklasifikasikan komentar pada sebuah teks bahasa Indonesia dan mengetahui seberapa besar akurasi yang didapatkan.

B. Proses Analisis Sentimen pada Dokumen

Langkah-langkah proses analisis metode sentiment:



Gambar 1 Proses Analisis Sentimen

C. Bahan yang digunakan

Data berupa teks sebanyak 100 komentar yang diambil dari akun resmi PT. Tiki Jalur Nugraha Eka Kurir (JNE) yang bersumber dari media sosial twitter.

D. Dokumen Twitter

Dalam table dibawah ini merupakan 10 contoh dokumen yang telah didapatkan peneliti dari twitter. Data ini telah peneliti klasifikasi secara manual sebagai contoh perhitungan manual:

Tabel 1 Contoh Klasifikasi

No	Text	Klasifikasi
1.	ngirim pake jne trucking lama banget dah	Negatif
2.	kami melayani kiriman anda sampai ke penjuru nusantara dan pelosok desa... percayakan pada JNE Juanda Depok saja yukk.	Positif
3.	@JNECare Iya parah tuh, punya saya juga dr jakarta kr bandung, dr hari sdh 8 hari gk nyampe	Negatif
4.	@JNECare Iya respon yanh menyatakan 3-6 hari? Padahal paket saya udah dari tanggal 5? Ini tanggalan JNE yang beda	Negatif
5.	Manfaatkan layanan jemput gratis dari JNE Kalimantan Blitar, Mudah dan Praktis.#JNEBLITAR	Positif
6.	Packing dirusak kek gini maksudnya apa ya? Mana barangnya diambil sendiri di gudang, dah 5hari ga diantar @JNE_ID	Negatif
7.	@Audioford1 selamat siang kak, maaf atas ketidaknyamanannya. terkait kendala yang terjadi dibantu tim @JNECare ya kak	Positif
8.	JNE lama dan tidak membantu , paket sudah 14 hari tidak bergerak tapi setiap hubungi CS selalu diminta tunggu	Negatif
9.	Bila tim pick up JNE Juanda Depok kurang berkenan atau kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti,terima kasih	Positif
10.	@JNECare tolong liat pesan saya ini email sy gk drespon ,paket saya salah alamat krn jne salah input alamat ini gmnaa!!	Negatif

E. Preprocessing

Tahap ini merupakan tahap membersihkan kata sebelum diolah lebih lanjut pada tahap berikutnya. Tahap preprocessing sangat penting, karena pada tahap ini semua perubahan kata akan memiliki nilai sendiri yang mempengaruhi akurasi system yang akan dibuat nantinya. Berikut tahapan Preprocessing:

1. Case Folding

Tahapan Case Folding merupakan sebuah proses yang digunakan untuk pengubahan teks yang berhuruf kapital menjadi huruf kecil dan hanya digunakan untuk huruf 'a' sampai dengan 'z' yang diterima. Selain karakter huruf dihilangkan dan dianggap pembatas.

Tabel 2 Proses Case Folding

No	Input	Output
1.	ngirim pake jne trucking lama banget dah	ngirim pake jne trucking lama banget dah
2.	kami melayani kiriman anda sampai ke penjuru nusantara dan pelosok desa... percayakan pada JNE Juanda Depok saja yukk.	kami melayani kiriman anda sampai ke penjuru nusantara dan pelosok desa... percayakan pada jne juanda depok saja yukk.
3.	@JNECare Iya parah tuh, punya saya juga dr jakarta kr bandung, dr hari sdh 8 hari gk nyampe	@jncare iya parah tuh, punya saya juga dr jakarta kr bandung, dr hari sdh 8 hari gk nyampe
4.	@JNECare Iya respon yanh menyatakan 3-6 hari? Padahal paket saya udah dari tanggal 5?	@jncare iya respon yanh menyatakan 3-6 hari? padahal paket saya udah dari tanggal 5? ini tanggalan

No	Input	Output
	Ini tanggalan JNE yang beda	jne yang beda
5.	Manfaatkan layanan jemput gratis dari JNE Kalimantan Blitar, Mudah dan Praktis.#JNEBLITAR	manfaatkan layanan jemput gratis dari jne kalimantan blitar, mudah dan praktis.#jneblitar
6.	Packing dirusak kek gini maksudnya apa ya? Mana barangnya diambil sendiri di gudang, dah 5hari ga diantar @JNE_ID	packing dirusak kek gini maksudnya apa ya? mana barangnya diambil sendiri di gudang, dah 5hari ga diantar @jne_id
7.	@Audioford1 selamat siang kak, maaf atas ketidaknyamanannya. terkait kendala yang terjadi dibantu tim @JNECare ya kak	@audioford1 selamat siang kak, maaf atas ketidaknyamanannya. terkait kendala yang terjadi dibantu tim @jncare ya kak
8.	JNE lama dan tidak membantu, paket sudah 14 hari tidak bergerak tapi setiap hubungi CS selalu diminta tunggu	jne lama dan tidak membantu, paket sudah 14 hari tidak bergerak tapi setiap hubungi cs selalu diminta tunggu
9.	Bila tim pick up JNE Juanda Depok kurang berkenan atau kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti,terima kasih	bila tim pick up jne juanda depok kurang berkenan atau kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti,terima kasih
10.	@JNECare tolong liat pesan saya ini email sy gk drespon.paket saya salah alamat krn jne salah input alamat ini gmnaa!!	@jncare tolong liat pesan saya ini email sy gk drespon.paket saya salah alamat krn jne salah input alamat ini gmnaa!!

2. Tokenizing

Tahapan tokenizing merupakan tahapan proses untuk menguraikan deskripsi yang semulanya berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter seperti tanda titik (.), koma (,) dan spasi pada karakter kata pada sebuah kalimat tersebut :

Tabel 3 Proses Tokenizing

No	Input	Output
1.	ngirim pake jne trucking lama banget dah	ngirim pake jne trucking lama banget dah
2.	kami melayani kiriman anda sampai ke penjurur nusantara dan pelosok desa... percayakan pada jne juanda depok saja yukk.	kami melayani kiriman anda sampai ke penjurur nusantara dan pelosok desa percayakan pada jne juanda depok saja yukk
3.	@jncare iya parah tuh, punya saya juga dr jakarta kr bandung, dr hari sdh 8 hari gk nyampe	jncare iya parah tuh punya saya juga dr jakarta kr bandung dr hari sdh 8 hari gk nyampe
4.	@jncare iya respon yanh menyatakan 3-6 hari? padahal paket saya udah dari tanggal 5? ini tanggalan jne yang beda	jncare iya respon yanh menyatakan 36 hari padahal paket saya udah dari tanggal 5 ini tanggalan jne yang beda
5.	manfaatkan layanan jemput gratis dari jne kalimantan blitar, mudah dan praktis.#jneblitar	manfaatkan layanan jemput gratis dari jne kalimantan blitar mudah dan praktis jneblitar
6.	packing dirusak kek gini maksudnya apa ya? mana barangnya diambil sendiri di gudang, dah 5hari ga diantar @jne_id	packing dirusak kek gini maksudnya apa ya mana barangnya diambil sendiri di gudang dah 5hari ga diantar jneid
7.	@audioford1 selamat siang kak, maaf atas ketidaknyamanannya. terkait kendala yang terjadi dibantu tim @jncare ya kak	audioford1 selamat siang kak maaf atas ketidaknyamanannya terkait kendala yang terjadi dibantu tim jncare ya kak
8.	jne lama dan tidak membantu, paket sudah 14 hari tidak bergerak tapi setiap hubungi cs selalu diminta tunggu	jne lama dan tidak membantu paket sudah 14 hari tidak bergerak tapi setiap hubungi cs selalu diminta tunggu
9.	bila tim pick up jne juanda depok kurang berkenan atau kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti,terima kasih	bila tim pick up jne juanda depok kurang berkenan atau kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti terima kasih
10.	@jncare tolong liat pesan saya ini email sy gk drespon.paket saya salah alamat krn jne salah input alamat ini gmnaa!!	jncare tolong liat pesan saya ini email sy gk dresponpaket saya salah alamat krn jne salah input alamat ini gmnaa

3. Filtering / Stopword Removal

Stopword Removal adalah proses menghilangkan kata yang tidak penting, guna merampingkan perhitungan terhadap suatu dokumen. Penghilangan kata pada tahap ini misalnya pada kata 'di', 'ada', 'ketika' dll. Karena dokumen pada twitter merupakan data yang sangat tidak terstruktur, sehingga untuk memproses bagian ini harus ditambahkan list kata tambahan. Berikut merupakan contoh table luaran proses stopword removal:

Tabel 4 Proses Filtering/ Stopword Removal

No	Input	Output
1.	ngirim pake jne trucking lama banget dah	ngirim pake jne trucking lama banget
2.	kami melayani kiriman anda sampai ke penjuru nusantara dan pelosok desa percayakan pada jne juanda depok saja yukk	kami melayani kiriman anda sampai penjuru nusantara pelosok desa percayakan jne juanda depok saja yukk
3.	jnecare iya parah tuh punya saya juga dr jakarta kr bandung, dr hari sdh 8 hari gk nyampe	jnecare parah tuh punya saya juga dr jakarta kr bandung dr hari sdh 8 hari gk nyampe
4.	jnecare iya respon yanh menyatakan 36 hari padahal paket saya udah dari tanggal 5 ini tanggalan jne yang beda	jnecare respon yanh menyatakan 36 hari padahal paket saya udah tanggal 5 tanggalan jne beda
5.	manfaatkan layanan jemput gratis dari jne kalimantan blitar mudah dan praktis jneblitar	manfaatkan layanan jemput gratis jne kalimantan blitar mudah praktis jneblitar
6.	packing dirusak kek gini maksudnya apa ya mana barangnya diambil sendiri di gudang dah Shari ga diantar jneid	packing dirusak kek gini maksudnya mana barangnya diambil sendiri gudang dah Shari ga diantar jneid
7.	audioford1 selamat siang kak maaf atas ketidaknyamanannya terkait kendala yang terjadi dibantu tim jnecare ya kak	audioford1 selamat siang kak maaf atas ketidaknyamanannya terkait kendala terjadi dibantu tim jnecare kak
8.	jne lama dan tidak membantu paket sudah 14 hari tidak bergerak tapi setiap hubungi cs selalu diminta tunggu	jne lama membantu paket sudah 14 hari bergerak tapi setiap hubungi cs selalu diminta tunggu
9.	bila tim pick up jne juanda depok kurang berkenan atau kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti,terima kasih	bila tim pick up jne juanda depok kurang berkenan kurang sopan dalam pelayanannya bisa mention kami utk ditindak lanjuti terima kasih
10.	jnecare tolong liat pesan saya ini email sy gk dresponpaket saya salah alamat krn jne salah input alamat ini gmnaa	jnecare tolong liat pesan saya email sy gk dresponpaket saya salah alamat krn jne salah input alamat gmnaa

4. Stemming

Proses Stemming merupakan mengubah kata yang berimbuhan menjadi kata dasar, pada tahap ini peneliti menggunakan library dari sastrawi untuk melakukan perubahan pada kata berikut:

Tabel 5 Proses Stemming

No	Input	Output
1.	ngirim pake jne trucking lama banget	ngirim pake jne trucking lama banget
2.	kami melayani kiriman anda sampai penjuru nusantara pelosok desa percayakan jne juanda depok saja yukk	kami layan kirim anda sampai penjuru nusantara pelosok desa percaya jne juanda depok saja yukk
3.	jnecare iya parah tuh punya saya juga dr jakarta kr bandung dr hari sdh 8 hari gk nyampe	jnecare parah tuh punya saya juga dr jakarta kr bandung dr hari sdh 8 hari gk nyampe
4.	jnecare iya respon yanh menyatakan 36 hari padahal paket saya udah tanggal 5 tanggalan jne beda	jnecare respon yanh nyata 36 hari padahal paket saya udah tanggal 5 tanggal jne beda
5.	manfaatkan layanan jemput gratis jne kalimantan blitar mudah praktis jneblitar	manfaat layan jemput gratis jne kalimantan blitar mudah praktis jneblitar
6.	packing dirusak kek gini maksudnya mana barangnya diambil sendiri gudang dah 5hari ga diantar jneid	packing rusak kek gini maksud mana barang ambil sendiri gudang dah 5hari ga antar jneid
7.	audioford1 selamat siang kak maaf atas ketidaknyamanannya terkait kendala terjadi	audioford1 selamat siang kak maaf atas tidak nyaman kait kendala jadi bantu tim jnecare kak

No	Input	Output
	dibantu tim jne care kak	
8.	jne lama membantu paket sudah 14 hari bergerak tapi setiap hubungi cs selalu diminta tunggu	jne lama bantu paket sudah 14 hari gerak tapi tiap hubungi cs selalu minta tunggu
9.	bila tim pick up jne juanda depok kurang berkenan kurang sopan dalam pelayanannya bisa mention kami utk tindak lanjut terima kasih	bila tim pick up jne juanda depok kurang kenan kurang sopan dalam layan bisa mention kami utk tindak lanjut terima kasih
10.	jne care tolong liat pesan saya email sy gk drespon paket saya salah alamat krn jne salah input alamat gmnaa	jne care tolong liat pesan saya email sy gk drespon paket saya salah alamat krn jne salah input alamat gmnaa

F. Pembobotan TF/ IDF

Pembobotan kata yang peneliti gunakan disini adalah pembobotan TF-IDF. Menentukan kata kunci terlebih dahulu sebelum melakukan perhitungan. Berikut ini merupakan contoh penerapan TF pada data testing di dokumen.

Tabel 6 Data Testing

D9	bila tim pick up jne juanda depok kurang kenan kurang sopan dalam layan bisa mention kami utk tindak lanjut terima kasih
----	--

1. Menentukan Nilai TF

Berikut ini merupakan bentuk pencarian TF pada data testing dengan mencari kata yang muncul dalam dokumen.

Tabel 7 Pencarian TF Pada Data Testing

bila	1
tim	1
pick	1
up	1
jne	1
juanda	1
depok	1
kurang	1
sopan	1
dalam	1
layan	1
bisa	1
mention	1
kami	1
utk	1
tindak	1
lanjut	1
terima	1
kasih	1

Berikut ini adalah pencarian nilai TF dan DF pada data training., nilai TF merupakan kata yang sering muncul pada dokumen sedangkan DF merupakan jumlah kata yang muncul dari dokumen. Di bawah ini contoh pencarian TF dan DF pada data training.

Tabel 8 Pencarian TF dan DF

NO	Term	TF										DF
		D9	D1	D2	D3	D4	D5	D6	D7	D8	D10	
1	bila	1	0	0	0	0	0	0	0	0	0	1
2	tim	1	0	0	0	0	0	0	0	0	0	1
3	pick	1	0	0	0	0	0	0	0	0	0	1
4	up	1	0	0	0	0	0	0	0	0	0	1
5	jne	1	1	1	0	0	1	0	0	1	1	6
6	juanda	1	0	1	0	0	0	0	0	0	0	2
7	depok	1	0	1	0	0	0	0	0	0	0	1
8	kurang	1	0	0	0	0	0	0	0	0	0	1
9	sopan	1	0	0	0	0	0	0	0	0	0	1
10	dalam	1	0	0	0	0	0	0	0	0	0	1
11	layan	1	0	1	0	0	1	0	0	0	0	3

NO	Term	TF										DF
		D9	D1	D2	D3	D4	D5	D6	D7	D8	D10	
12	bisa	1	0	0	0	0	0	0	0	0	0	1
13	mention	1	0	0	0	0	0	0	0	0	0	1
14	kami	1	0	1	0	0	0	0	0	0	0	2
15	utk	1	0	0	0	0	0	0	0	0	0	1
16	tindak	1	0	0	0	0	0	0	0	0	0	1
17	lanjut	1	0	0	0	0	0	0	0	0	0	1
18	terima	1	0	0	0	0	0	0	0	0	0	1
19	kasih	1	0	0	0	0	0	0	0	0	0	1

Terdapat sebagai contoh kata “dapet” muncul 2 kali pada D2 berarti TF D2 = 2, sedangkan nilai DF = 1 di karenakan DF hanya akan menghitung jumlah kata yang mewakili setiap dokumen, buatkan menghitung jumlah kata yang muncul dalam dokumen.

2. Menghitung nilai IDF

Proses perhitungan idf menggunakan rumus (2.1). Dimana D merupakan jumlah dokumen dan DF merupakan dokumen frekuensi atau jumlah kata yang akan muncul pada setiap dokumen. Untuk mencari IDF bisa menggunakan rumus $\text{Log}(D/DF)$. Contoh pada kata “terima” yang memiliki adalah D9 dan DF = 1. Maka bisa dihitung seperti berikut :

$$\begin{aligned}
 \text{Misalkan IDF bayar} &= \text{Log} \left(\frac{D}{DF} \right) \\
 &= \text{Log} \left(\frac{10}{1} \right) \\
 &= \text{Log } 10 = 1.000
 \end{aligned}$$

Tabel 9 Pencarian IDF

NO	Term	TF										DF	IDF
		D9	D1	D2	D3	D4	D5	D6	D7	D8	D10		
1	bila	1	0	0	0	0	0	0	0	0	0	1	1,000
2	tim	1	0	0	0	0	0	0	0	0	0	1	1,000
3	pick	1	0	0	0	0	0	0	0	0	0	1	1,000
4	up	1	0	0	0	0	0	0	0	0	0	1	1,000
5	jne	1	1	1	0	0	1	0	0	1	1	6	0,222
6	juanda	1	0	1	0	0	0	0	0	0	0	2	0,699
7	depok	1	0	1	0	0	0	0	0	0	0	1	1,000
8	kurang	1	0	0	0	0	0	0	0	0	0	1	1,000
9	sopan	1	0	0	0	0	0	0	0	0	0	1	1,000
10	dalam	1	0	0	0	0	0	0	0	0	0	1	1,000
11	layan	1	0	1	0	0	1	0	0	0	0	3	0,523
12	bisa	1	0	0	0	0	0	0	0	0	0	1	1,000
13	mention	1	0	0	0	0	0	0	0	0	0	1	1,000
14	kami	1	0	1	0	0	0	0	0	0	0	2	0,699
15	utk	1	0	0	0	0	0	0	0	0	0	1	1,000
16	tindak	1	0	0	0	0	0	0	0	0	0	1	1,000
17	lanjut	1	0	0	0	0	0	0	0	0	0	1	1,000
18	terima	1	0	0	0	0	0	0	0	0	0	1	1,000
19	kasih	1	0	0	0	0	0	0	0	0	0	1	1,000

3. Mencari TF IDF

Proses ini pencarian nilai tf idf dengan mengalikan nilai tf dengan nilai idf.

Tabel 10 Pencarian TF IDF

wdt=tf.idf									
D9	D1	D2	D3	D4	D5	D6	D7	D8	D10
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
0,222	0,22184875	0,22184875	0	0	0,22184875	0	0	0,22184875	0,22184875
0,699	0	0,698970004	0	0	0	0	0	0	0

wdt=tf.idf									
D9	D1	D2	D3	D4	D5	D6	D7	D8	D10
1,000	0	1,000	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
0,523	0	0,523	0	0	0,522878745	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
0,699	0	0,699	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0

Berdasarkan tabel di atas untuk menentukan nilai tf idf dengan cara mengalikan nilai tf dengan dengan nilai idf. Berikut ini sebagai contoh:

$$\begin{aligned} \text{Bobot 'kami' (D2)} &= \text{TF} \times \text{IDF} \\ &= 1 \times 0,699 \\ &= 0,699 \end{aligned}$$

G. Tahap Klasifikasi

Tahap ini merupakan tahap lanjutan dari tahap pembobotan yaitu tahapan nilai similaritas atau nilai kemiripan antara data testing dengan data training menggunakan algoritma KNN. Algoritma ini harus menentukan nilai k dan

mengambil nilai kemiripan yang paling tinggi. Kemudian menyesuaikan dengan hasil kemiripan yang sesuai dengan nilai k. Berikut ini contoh perkalian data testing dengan data training:

Tabel 11 Perkalian jumlah data testing dengan data training

<i>WD9*Wdi</i>								
D1	D2	D3	D4	D5	D6	D7	D8	D10
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0,049217	0,049217	0	0	0,049217	0	0	0,049217	0,049217
0	0,488559	0	0	0	0	0	0	0
0	1,000	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0,273	0	0	0,273402	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0,489	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0,0492	2,2997	0,0000	0,0000	0,3226	0,0000	0,0000	0,0492	0,0492

Tahap selanjutnya yaitu mencari panjang vektor dengan cara mengkalikan akar jumlah data testing sudah dikudratkan dan akar jumlah data training yang sudah di kuadratkan.

Tabel 12 Penjumlahan Vektor

Panjang vector									
D9	D1	D2	D3	D4	D5	D6	D7	D8	D10
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
0,049	0,049217	0,049217	0	0	0,049217	0	0	0,049217	0,049217

Panjang vector									
D9	D1	D2	D3	D4	D5	D6	D7	D8	D10
0,489	0	0,488559	0	0	0	0	0	0	0
1,000	0	1,000	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
0,273	0	0,273	0	0	0,273402	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
0,489	0	0,489	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0	0	0
16,2997	0,0492	2,2997	0,0000	0,0000	0,3226	0,0000	0,0000	0,0492	0,0492
4,0373	0,2218	1,5165	0,0000	0,0000	0,5680	0,0000	0,0000	0,2218	0,2218

Selanjutnya yaitu mencari nilai jarak kemiripan, seperti contoh dibawah ini:

Tabel 13 Hasil Consine Similarity

Hasil Perhitungan Tabel								
D1	D2	D3	D4	D5	D6	D7	D8	D10
0,0549	0,3756	0	0	0,140687	0	0	0,05495	0

Berdasarkan tabel diatas untuk perhitungan Consine Similarity bisa dilihat seperti contoh di bawah ini:

$$\begin{aligned} \text{Cosine}(D_1) &= \frac{\sum(Q \times D_9)}{\sqrt{Q^2 \times D_9^2}} \\ &= 0.0492 \quad (4.0373 \times 0.2218) \\ &= 0.0549 \end{aligned}$$

Setelah perhitungan Consine Similarity selesai, selanjutnya mengurutkan dokumen dari nilai terbesar ke terkecil. lalu menentukan nilai k . Disini peneliti menggunakan nilai $k = 3$ nilai yang paling tinggi kemiripan dengan data testing, berikut penjelasan hasil klasifikasi:

Tabel 14 Hasil Klasifikasi

Hasil Klasifikasi									
1	2	3	4	6	7	8	9	10	
D2	D5	D8	D1						

Dengan hasil kemiripan tersebut dapat di klasifikasikan bahwa D9 termasuk kelas (positif). Berikut ini percobaan dengan menambahkan nilai k dengan batasan angka ganjil.

Jika $k = 1$, maka D9 masuk ke kelas C1 (positif), diwakili oleh dokumen D2

Jika $k = 3$, maka D9 masuk ke kelas C2 (negatif), diwakili oleh dokumen D2, D5 dan D8

Jika $k=5$, maka D9 masuk ke kelas C1 (positif), diwakili oleh dokumen D2,D5,D8, dan D1

III. HASIL DAN PEMBAHASAN

1. Pengambilan Data dari API Twitter

Proses pengambilan data dari API Twitter sekaligus menampilkan hasil sentimen dari komentar netizen.

index	Tanggal	Pengguna	Tweet	Sentimen
0	2023-03-01 03:13:57	harisNO	_uncover_ID_ i love , cuma masih bnyk seller yg malah nolak pake pos. glau kenapa	Positif
1	2023-03-01 03:13:48	Blifromdahood	RT_uncover: hallo_ID saya kirim hp oppo di ganti nokia, memang tidak ada video buka paket hp. bapak saya gaktek, bapa...	Netral
2	2023-03-01 03:13:29	Prahaya_Gilang	_Anton_miring_uncover_ID_ Kalau ini jangan bang hehehe Tidak aaminn	Netral
3	2023-03-01 03:13:16	faridassutisna	RT_uncover: hallo_ID saya kirim hp oppo di ganti nokia, memang tidak ada video buka paket hp. bapak saya gaktek, bapa...	Netral
4	2023-03-01 03:12:57	JNEDEPOK_ID	bingung mau kirim paket barang / dokumen kemana ? Ke JNE Depok aja . Kami siap melayani Anda .)	Positif
5	2023-03-01 03:12:56	ratiVng	RT_uncover: hallo_ID saya kirim hp oppo di ganti nokia, memang tidak ada video buka paket hp. bapak saya gaktek, bapa...	Netral
6	2023-03-01 03:12:05	AndiParuli	RT_uncover: hallo_ID saya kirim hp oppo di ganti nokia, memang tidak ada video buka paket hp. bapak saya gaktek, bapa...	Netral
7	2023-03-01 03:11:43	happieszperson	RT_uncover: hallo_ID saya kirim hp oppo di ganti nokia, memang tidak ada video buka paket hp. bapak saya gaktek, bapa...	Netral
8	2023-03-01 03:11:28	agumwicaksana	RT_uncover: hallo_ID saya kirim hp oppo di ganti nokia, memang tidak ada video buka paket hp. bapak saya gaktek, bapa...	Netral
9	2023-03-01 03:11:14	AnungLuhung	_ID Kandani mending nggo si merah penguasa tajam kok, leres mboten pakde 🤔	Netral
10	2023-03-01 03:10:37	JNE_ID	Masih Kak, yuk langsung diceki 🤔	Netral

Gambar 2 Data API Twitter

2. Penambahan kolom untuk pelabelan

Penambahan kolom untuk pelabelan difungsikan untuk mempermudah mengkategorikan positif, netral, negatif.

Unnamed: 0	Tanggal	Pengguna	Tweet	Sentimen	Label
0	2023-03-01 03:13:57	harisNO	_uncover_ID_ i love , cuma masih bnyk seller...	Positif	1
1	2023-03-01 03:13:48	Blifromdahood	RT_uncover: hallo_ID saya kirim hp oppo di g...	Netral	0
2	2023-03-01 03:13:29	Prahaya_Gilang	_Anton_miring_uncover_ID_ Kalau ini jangan...	Netral	0
3	2023-03-01 03:13:16	faridassutisna	RT_uncover: hallo_ID saya kirim hp oppo di g...	Netral	0
4	2023-03-01 03:12:57	JNEDEPOK_ID	bingung mau kirim paket barang / dokumen kema...	Positif	1
5	2023-03-01 03:12:56	ratiVng	RT_uncover: hallo_ID saya kirim hp oppo di g...	Netral	0
6	2023-03-01 03:12:05	AndiParuli	RT_uncover: hallo_ID saya kirim hp oppo di g...	Netral	0
7	2023-03-01 03:11:43	happieszperson	RT_uncover: hallo_ID saya kirim hp oppo di g...	Netral	0
8	2023-03-01 03:11:28	agumwicaksana	RT_uncover: hallo_ID saya kirim hp oppo di g...	Netral	0
9	2023-03-01 03:11:14	AnungLuhung	_ID Kandani mending nggo si merah penguasa t...	Netral	0
10	2023-03-01 03:10:37	JNE_ID	Masih Kak, yuk langsung diceki 🤔	Netral	0

Gambar 3 Penambahan Kolom

6. Tokenizing

Tahapan tokenizing merupakan tahapan proses untuk menguraikan deskripsi yang awalnya berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter seperti underscore (_), titik dua (:), dan spasi pada karakter kata pada sebuah kalimat.

```

def identify_tekens(row) :
    text = row['Tweet']
    tokens = nltk.word_tokenize(text)
    token_words = [w for w in tokens if w.isalpha()]
    return token_words

csv_data['Tweet'] = csv_data.apply(identify_tekens, axis=1)
csv_data.Tweet

0      [uncover, id, i, love, cuma, masih, bnyk, sell...
1      [rt, uncover, hallo, id, saya, kirim, hp, oppo...
2      [anton, miring, uncover, id, kalau, ini, janga...
3      [rt, uncover, hallo, id, saya, kirim, hp, oppo...
4      [bingung, mau, kirim, paket, barang, dokumen, ...
...
95     [rt, uncover, hallo, id, saya, kirim, hp, oppo...
96     [rt, uncover, hallo, id, saya, kirim, hp, oppo...
97     [rt, hallo, id, saya, kirim, hp, oppo, di, gan...
98     [rt, uncover, hallo, id, saya, kirim, hp, oppo...
99     [rt, uncover, hallo, id, saya, kirim, hp, oppo...
Name: Tweet, Length: 100, dtype: object

```

Gambar 7 Tokenizing

7. Stemming

Stemming mengubah kata yang berlebihan menjadi kata dasar, pada tahap ini menggunakan library dari sastrawi untuk melakukan perubahan kalimat.

```

def stem_list(row) :
    text = row['Tweet']
    stem = [stemming.stem(word) for word in text]
    return(stem)

csv_data['Tweet'] = csv_data.apply(stem_list, axis=1)
csv_data.Tweet

0      [uncov, id, i, love, cuma, masih, bnyk, seller...
1      [rt, uncov, hallo, id, saya, kirim, hp, oppo, ...
2      [anton, mire, uncov, id, kalau, ini, jangan, b...
3      [rt, uncov, hallo, id, saya, kirim, hp, oppo, ...
4      [bingung, mau, kirim, paket, barang, dokumen, ...
...
95     [rt, uncov, hallo, id, saya, kirim, hp, oppo, ...
96     [rt, uncov, hallo, id, saya, kirim, hp, oppo, ...
97     [rt, hallo, id, saya, kirim, hp, oppo, di, gan...
98     [rt, uncov, hallo, id, saya, kirim, hp, oppo, ...
99     [rt, uncov, hallo, id, saya, kirim, hp, oppo, ...
Name: Tweet, Length: 100, dtype: object

```

Gambar 8 Stemming

8. Stopword

Proses menghilangkan kata yang tidak penting, guna merampingkan perhitungan terhadap suatu dokumen.

Unnamed: 0	Tweet	label
0	[uncov, id, i, love, cuma, masih, bnyk, seller...	1
1	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
2	[anton, mire, uncov, id, kalau, ini, jangan, b...	0
3	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
4	[bingung, mau, kirim, paket, barang, dokumen, ...	1
5	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
6	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
7	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
8	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
9	[id, kandani, mend, nggo, si, merah, penguasa,...	0
10	[masih, kak, yuk, langsung, dicek]	0
11	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
12	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
13	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0
14	[rt, uncov, hallo, id, saya, kirim, hp, oppo, ...	0

Gambar 9 Stopword

9. Menghitung TF-IDF

Proses menghitung jumlah kata yang akan muncul pada setiap dokumen.

TF-IDF Vectorizer														
	0	1	2	3	4	5	6	7	8	9	...	429	430	431 \
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0
...
96	0	0	0	0	0	0	0	0	0	0	...	0	0	0
97	0	0	0	0	0	0	0	0	0	0	...	0	0	0
98	0	0	0	0	0	0	0	0	0	0	...	0	0	0
99	0	0	0	0	0	0	0	0	0	0	...	0	0	0
100	0	0	0	0	0	0	0	0	0	0	...	0	0	0
	432	433	434	435	436	437	438							
1	0	0	0	0	0	1	0							
2	0	0	0	0	0	0	0							
3	0	0	0	0	0	0	0							
4	0	0	0	0	0	0	0							
5	0	0	0	0	0	0	0							
...							
96	0	0	0	0	0	0	0							
97	0	0	0	0	0	0	0							
98	0	0	0	0	0	0	0							
99	0	0	0	0	0	0	0							
100	0	0	0	0	0	0	0							

Gambar 10 Perhitungan TF-IDF

10. Klasifikasi

```

✓ 0s ▶ from sklearn.metrics import classification_report, confusion_matrix
cm = confusion_matrix(y_test, y_pred)
cr = classification_report(y_test, y_pred)
print(cm)
print(cr)

```

```

[[18  0]
 [ 2  0]]

```

	precision	recall	f1-score	support
0	0.90	1.00	0.95	18
1	0.00	0.00	0.00	2
accuracy			0.90	20
macro avg	0.45	0.50	0.47	20
weighted avg	0.81	0.90	0.85	20

Gambar 11 Klasifikasi

IV. KESIMPULAN

Perhitungan algoritma K-NN menghasilkan klasifikasi yang diimplementasikan terhadap data twitter sebanyak 100 komentar dengan persentase akurasi sebesar 90%, dengan nilai yang diuji 20 data (20%) sehingga dapat disimpulkan bahwa hasil klasifikasi ini dapat digunakan untuk penilaian pelayanan JNE dari masyarakat umum melalui media sosial Twitter.

UCAPAN TERIMA KASIH

Saya mengucapkan terimakasih kepada Kedua orangtua saya yang selalu mendoakan dan memotivasi saya yang selalu memberikan semangat, Rektor Universitas Muhammadiyah Sidoarjo, Bapak Dr. Hidayatulloh, M.Si., para pimpinan, staf, dan seluruh civitas akademika Universitas Muhammadiyah Sidoarjo, Dekan Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo, Bapak Dr. Hindarto, S.Kom, MT. Serta Kepala Program Studi Informatika, Bapak Ir. Sumarno MM., Dosen Pembimbing, Bapak Hamzah Setiawan, S.Kom., M.Kom. yang telah banyak membimbing dan membantu dalam penelitian skripsi ini, Serta semua pihak yang tidak dapat sebutkan satu persatu, yang telah ikhlas memberikan doa serta motivasi.

REFERENSI

- [1] K. Murah, "Kargo Murah," Sabtu September 2021. [Online]. Available: <https://www.kargomurah.co.id/apa-itu-jne-dan-layanannya/#:~:text=JNE%20merupakan%20singkatan%20dari%20PT,Soeprpto%20Suparno%20pada%20tahun%201990..>
- [2] I. Sebastian, "Edukasi," Jum'at Januari 2015. [Online]. Available: <https://edukasi.okezone.com/read/2014/12/31/65/1086393/ramai-ramai-mengkritik-di-media-sosial>.
- [3] V. Elvira, "industri," Kamis Juli 2022. [Online]. Available: <https://industri.kontan.co.id/news/jne-proyeksikan-volume-pengiriman-paket-tumbuh-40-hingga-akhir-tahun>.
- [4] wikipedia, "id," 30 Maret 2022. [Online]. Available: <https://id.wikipedia.org/wiki/JNE>.
- [5] wikipedia, "id," 19 Juli 2022. [Online]. Available: <https://id.wikipedia.org/wiki/Twitter>.
- [6] M. A. Ramdhani and O. N. Rahim, "Analisis sentimen untuk mengukur popularitas tokoh publik berdasar data pada media sosial twitter menggunakan algoritma data mining dengan teknik klasifikasi," *Informasi*, vol. VI, no. 2, pp. 1–15, 2014.
- [7] E. Helmud, "Optimasi Basis Data Oracle Menggunakan Complex View Studi Kasus : PT. Berkat Optimis Sejahtera (PT.BOS) Pangkalpinang," *J. Informanika*, vol. 7, no. 1, pp. 80–86, 2021.
- [8] Q. Budiman, S. Mouton, L. Veenhoff, and A. Boersma, "程威特 1, 吴海涛 1, 江帆 2," *J. Inov. Penelit.*, vol. 1, no. 0.1101/2021.02.25.432866, pp. 1–15, 2021.

-
- [9] Muhammad Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *JTIM J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2020.
- [10] Hartanto, "Hartanto 2017 text mining dan sentimen analisis twitter," *J. Psikol. Ilm.*, vol. 9, no. 1, pp. 18–25, 2017.
- [11] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.

Jurnal_Ricky_Renaldo_A_161080200182.docx

ORIGINALITY REPORT

7%

SIMILARITY INDEX

7%

INTERNET SOURCES

0%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1 repositori.usu.ac.id 2%
Internet Source

2 ejurnal.stmik-budidarma.ac.id 2%
Internet Source

3 www.is.its.ac.id 2%
Internet Source

4 media.neliti.com 2%
Internet Source

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On