

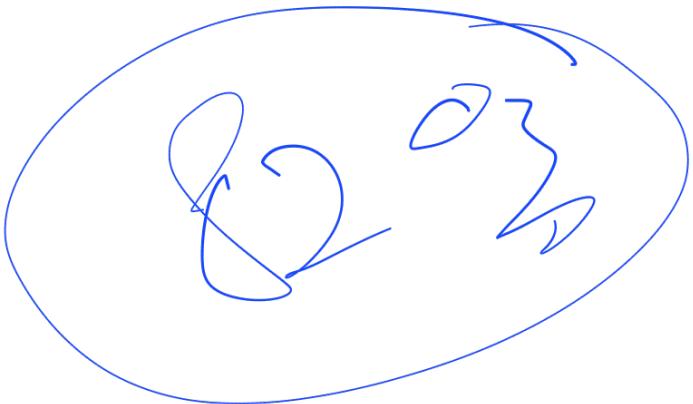
1 : 8 0%

2 : 3 5%

3 : 1 5%

4 : 1 5%

5 : 5%



SIS_SENTIMENT_JASA_EKSPEDIS
SICEPAT_-
_Ferren_Hillary_Kalalo.pdf
by

Submission date: 15-May-2022 06:30PM (UTC+0700)

Submission ID: 1836577306

File name: SIS_SENTIMENT_JASA_EKSPEDISI_SICEPAT_-_Ferren_Hillary_Kalalo.pdf (643.42K)

Word count: 4567

Character count: 28763

**MEMBANDINGKAN KINERJA ALGORITMA SUPPORT VECTOR
MACHINE DAN RECURRENT NEURAL NETWORK TERHADAP**

~~**ANALISIS SENTIMENT JASA EKSPEDISI SICEPAT**~~

PROPOSAL SKRIPSI

Diajukan kepada

Fakultas Ilmu Komputer Universitas Klabat

Untuk Memenuhi Tuntutan Memperoleh Gelar Sarjana Komputer (S-1)

Program Studi Sistem Informasi



NIM : 105011910001

KALALO FERREN HILLARY

NIM : 105011910051

PANGALILA ABELARD JOVI JOHN

FAKULTAS ILMU KOMPUTER

UNIVERSITAS KLABAT

APRIL 2022

DAFTAR ISI

DAFTAR ISI	2
BAB I PENDAHULUAN	3
1.1 Latar Belakang Masalah	3
1.2 Rumusan Masalah	5
1.4 Manfaat Penelitian	6
1.5 Ruang Lingkup dan Batasan Penelitian	6
1.6 Daftar Istilah	7
3 BAB II TINJAUAN LITERATUR	8
2.1 Tinjauan Pustaka	8
2.2 Kajian Teori	10
2.2.1 Jasa Ekspedisi	10
2.2.2 Text Mining	10
2.2.3 Sentimen Analisis	12
2.2.4 Recurrent Neural Network	12
2.2.5 Support Vector Machine	13
2.2.6 Preprocessing	14
2.2.7 Confusion Matrix	16
2.3 Kerangka Konseptual	17
2.4 Hipotesis Penelitian	18
3 BAB III METODE PENELITIAN	19
3.1 Objek dan Subjek Penelitian	19
3.2 Populasi dan Sampel Penelitian	19
3.3 Metode Pengumpulan Data	20
3.4 Jenis dan Sumber Data	20
3.5 Variabel Penelitian	20
3.6 Metode Analisis Data	21
3.7 Tahapan Penelitian	21
DAFTAR PUSTAKA	27

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Seiring dengan berkembangnya zaman, maka kita sudah tidak asing lagi dengan yang namanya *online shopping* atau berbelanja lewat internet. Dengan berbelanja secara *online*, maka kita tidak perlu repot-repot lagi untuk pergi ke suatu tempat atau pergi ke suatu toko untuk sekedar berbelanja, karena saat ini kita sudah bisa membeli sesuatu tanpa perlu keluar rumah. Dapat dikatakan bahwa *online shopping* dan jasa ekspedisi merupakan satu paket komplit. Karena dengan berbelanja secara *online* kita juga memerlukan jasa yang dapat mengantarkan barang pesanan tersebut ke alamat kita. Tentunya baik setiap *online shop* maupun *customer* memiliki jasa ekspedisi favoritnya masing-masing, baik itu yang dipengaruhi oleh faktor biaya, pelayanan, merek atau *brand*, kualitas, iklan, jumlah cabang, ataupun dari persepsi setiap individu [1].

Pada bulan Juli 2021, ketua umum dari Asosiasi Logistik Indonesia (ALI) memaparkan bahwa dalam dunia logistik, ada 3 tahapan pengantaran. Pertama yaitu *first-mile delivery*, yang merupakan pengiriman bahan baku ke pihak pabrik untuk menunjang proses produksi barang. Kedua yaitu *mid-mile delivery*, yang merupakan pengiriman produk jadi dari pabrik ke toko. Lalu yang ketiga yaitu *last-mile delivery*, yang mana merupakan pengiriman barang dari toko ke konsumen. Dan selama masa pandemi Covid-19 pertumbuhan arus pengiriman pada tahapan *last-mile delivery* mengalami peningkatan sebesar 30 - 40% adalah. [2]

Belakangan ini ada salah satu jasa ekspedisi di Indonesia yang tengah menjadi topik hangat yang sering diperbincangkan oleh masyarakat Indonesia lewat beberapa *platform* media sosial seperti Twitter dan Tiktok. Jasa ekspedisi tersebut adalah SiCepat. SiCepat merupakan jasa ekspedisi yang didirikan pada Tahun 2004, oleh seorang pria yang bernama Rudy Darwin Swigo. Untuk mencari tahu seperti apa

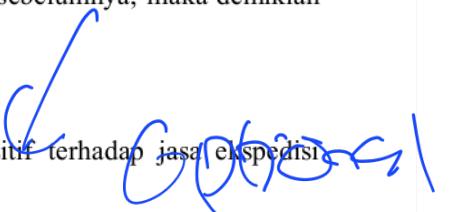
sebenarnya opini masyarakat Indonesia mengenai jasa ekspedisi SiCepat, tentunya kita harus bisa mencari informasi baik dengan cara mendengarkan ataupun membaca komentar-komentar yang dilontarkan oleh para *netizen* di dunia maya. Akan tetapi ada kendala yang mungkin dihadapi saat kita hendak melakukan hal tersebut, yaitu ada begitu banyak komentar-komentar yang tersedia, dan untuk membacanya satu persatu akan memerlukan waktu dan tenaga yang cukup besar.

Maka dari itu diperlukan suatu metode yang dapat dipakai untuk membantu mendapatkan informasi atas tanggapan atau komentar secara efisien serta efektif, dan salah satu cara yang dapat digunakan adalah dengan memanfaatkan yang namanya *Text Mining*, yang mana salah satu teknik analisisnya adalah sentimen analisis [3]. Sentimen analisis adalah suatu proses untuk memahami dan mengelola data berbentuk teks secara otomatis dengan tujuan untuk mendapatkan informasi mengenai sentimen yang ada di dalam suatu kalimat pendapat [4]. Tugas dari sentimen analisis adalah untuk mengelompokan teks dalam suatu kalimat berdasarkan sifat positif, negatif, ataupun netral [5]. Karena itu untuk mencari tahu bagaimana pandangan masyarakat Indonesia mengenai jasa ekspedisi siCepat, kita dapat menggunakan teknik sentimen analisis ini.

Karena sentimen analisis memiliki fungsi untuk menganalisis dan mengklasifikasikan kalimat, maka kita perlu membangunnya dengan menggunakan algoritma yang baik dalam bidang klasifikasi. Ada beberapa algoritma yang cukup sering digunakan dalam menganalisis sentimen, yaitu algoritma Naive Bayes Classifier, Support Vector Machine, Recurrent Neural Network, K-Nearest Neighbor, dan beberapa algoritma lainnya. Namun pada penelitian kali ini, peneliti akan berfokus untuk menganalisis sentimen atau opini masyarakat Indonesia mengenai jasa ekspedisi SiCepat dengan mengambil sumber data pada *platform* media sosial Twitter dengan menggunakan algoritma Support Vector Machine dan Recurrent Neural Network, kemudian peneliti juga akan turut membandingkan kinerja dari kedua algoritma tersebut.

1.2 Rumusan Masalah

Didasarkan pada latar belakang yang telah dijelaskan pada bagian yang sebelumnya, maka demikian rumusan masalah yang ada pada penelitian kali ini adalah:

1. Apakah masyarakat Indonesia memiliki opini atau sentimen yang positif terhadap jasa ekspedisi SiCepat? 
2. Algoritma manakah yang menghasilkan tingkat akurasi yang lebih tinggi antara Support Vector Machine dan Recurrent Neural Network saat menganalisis sentimen terhadap jasa ekspedisi SiCepat? 

1.3 Tujuan Penelitian

Tujuan utama dari penelitian kali ini adalah melakukan pengkategorian dari data yang diambil pada platform Twitter mengenai SiCepat, apakah memiliki sentimen yang positif, negatif, atau netral. Dan untuk mencapai hal itu peneliti memilih untuk menggunakan algoritma Support Vector Machine dan Recurrent Neural Network. Lalu setelah hasil dari kedua algoritma tersebut terlihat, maka akan dilakukan perbandingan tingkat akurasi, agar dapat mengetahui algoritma seperti apa yang lebih cocok atau lebih tepat untuk dipakai pada penelitian seperti ini.

1.4 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan kita bisa untuk melihat bagaimana pandangan masyarakat Indonesia terhadap jasa ekspedisi SiCepat. Apakah jasa ekspedisi SiCepat memiliki tanggapan positif atau negatif di mata masyarakat Indonesia, dilihat dari *tweet* yang beredar pada *platform* Twitter. Dan dengan dilakukannya perbandingan algoritma klasifikasi yang digunakan, maka kita bisa melihat algoritma mana yang memiliki tingkat akurasi yang lebih tinggi, dan lebih cocok untuk digunakan dalam teknik sentimen analisis seperti pada penelitian kali ini.

1.5 Ruang Lingkup dan Batasan Penelitian

Adapun ruang lingkup serta batasan penelitian yang terjadi pada penelitian kali ini yaitu:

1. Penelitian kali ini hanya mengambil sampel data pada *platform* Twitter saja.
2. Algoritma yang dibandingkan hanyalah algoritma Support Vector Machine dan Recurrent Neural Network yang sudah ada, tanpa melakukan perubahan pada parameter yang ada.
3. Penelitian kali ini bertipe *aspect-based sentiment analysis*.
4. Sampel data yang diambil akan terlebih dahulu dilakukan tahap *pre-processing* data sebelum diproses.
5. Data yang diambil hanyalah *tweet* pada tahun 2020 - 2022

1.6 Daftar Istilah

Online Shopping: proses dimana konsumen dapat secara langsung membeli sesuatu dari seorang penjual secara interaktif dan juga real-time melalui internet

Netizen: warganet atau warga internet, yaitu orang-orang yang secara aktif menggunakan teknologi internet

Dunia maya: suatu ruang informasi serta komunikasi di dalam internet

Algoritma: kumpulan instruksi atau langkah-langkah sistematis yang dipakai untuk memecahkan masalah matematis

Parameter: ukuran suatu keadaan secara relatif

Aspect-based sentiment analysis: tipe sentimen analisis yang digunakan untuk mengetahui aspek apa yang mendapatkan penilaian positif, negatif, ataupun netral

Pre-processing data: mengubah data mentah menjadi data yang dapat lebih mudah untuk dipahami serta digunakan

Sentimen: pendapat berdasarkan perasaan yang berlebih terhadap sesuatu

Klasifikasi: pengelompokan menurut standar yang telah ditetapkan

Text Mining: proses untuk mengekstrak informasi dari data yang ada dengan menganalisis pola-pola yang terbentuk

Delimiter : Delimiter adalah karakter khusus, biasanya berupa koma atau spasi, yang dipakai untuk memisahkan nama atau item dalam daftar.

BAB II TINJAUAN LITERATUR

2.1 Tinjauan Pustaka

Sebuah penelitian membutuhkan dukungan penelitian - penelitian sebelumnya yang berkaitan dengan penelitian yang dilakukan sekarang ini. Berikut ini adalah beberapa hasil penelitian yang pernah dilakukan berkaitan dengan penelitian ini :

NO	Peneliti	Objek	Metode	Hasil
1	Mohammad AL-Smadi,, Omar Qawasmehb, Mahmoud Al-Ayyouba, Yaser Jararweha, Brij Guptac	Ujicoba tingkat akurasi algoritma pada <i>Arabic hotels</i>	SVM, RNN	Hasilnya menunjukan SVM mengungguli RNN. tetapi waktu eksekusi RNN lebih unggul dibandingkan SVM.
2	Kurnia Muludi, Mohammad Surya Akbar, Dewi Asiah Shofiana, Admi Syarif	<i>Energy Independence Tweets</i>	RNN, Naive bayes, LSTM	Hasil akurasinya adalah Recurrent Neural Network 78%, Bernoulli Naive Bayes 67%, dan LSTM 75%.
3	Merinda Lestandy, Abdurrahim Abdurrahim, Lailis Syafa'ah	<i>Tweet vaksin covid-19</i>	RNN, Naive Bayes	Hasilnya RNN 97,77% dan Naive Bayes hanya sebesar 80%.
4	Sharazita Dyah Anggita, Ikmah	Komparasi	SVM,	Hasil dari pengujian yang

		Algoritma Klasifikasi Berbasis Particle Swarm Optimization Pada Analisis Sentimen Ekspedisi Barang	Naive Bayes	dilakukan mampu menghasilkan peningkatan akurasi sebesar 15.11% pada penerapan PSO di algoritma Naive Bayes. Peningkatan akurasi pada algoritma SVM berbasis PSO senilai 1.74% pada kernel sigmoid.
5	Nadia Ristya Dewi	Vaksinasi COVID-19	RNN	Hasilnya RNN-Word2Vec menghasilkan akurasi sebesar 51.71%, sedangkan RNN-Word2Vec menghasilkan akurasi sebesar 50.73%.
6	Auliya Rahman Isnain, Adam Indra Sakti, Debby Alita, Nurman Satya Marga	Kebijakan <i>lockdown</i> pemerintah jakarta	SVM	Hasil dari penelitian ini adalah accuracy sebesar 74% , precision sebesar 75%, recall sebesar 92% dan F1-Score sebesar 83%.

2.2 Kajian Teori

2.2.1 Jasa Ekspedisi

Jasa ekspedisi sangat erat kaitanya dengan proses kirim mengirim barang yang dilakukan oleh perusahaan penyedia jasa ekspedisi. Pesatnya perkembangan teknologi, usaha online atau perdagangan lainnya, maka perusahaan ekspedisi pun semakin beragam.

Jenis layanan yang ditawarkannya pun demikian, mulai menurut pengiriman cepat, ongkos yg murah dan penerapan tagihan pertanggungan dalam proses pengiriman. Jadi jasa ekspedisi merupakan perusahaan yg berkecimpung pada bidang pengiriman paket barang dimana setiap pelanggan diharuskan membayar ongkos kirim yang sesuai wilayah tujuan. Pengiriman dilakukan baik melalui jalur darat, laut dan udara ke semua wilayah Indonesia atau ke luar Indonesia.

2.2.2 Text Mining¹

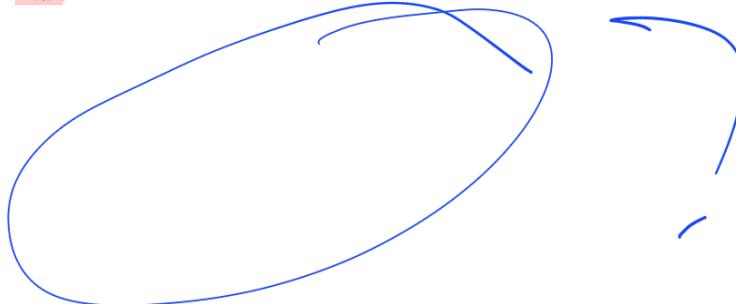
Text mining adalah bagian dari data mining, yaitu merupakan proses untuk mendapatkan suatu pengetahuan menggunakan alat analisis dimana pengguna berinteraksi dengan sekumpulan dokumen dari masa ² ke masa. Mirip seperti data mining, text mining berusaha untuk mengekstraksi informasi dari suatu sumber data (kumpulan dokumen) melalui identifikasi dan eksplorasi pola yang ada. Terdapat beberapa ⁴ area penerapan *text mining*, yaitu:

1. **Ekstraksi Informasi (Information Extraction)** Identifikasi frasa kunci dan keterikatan di dalam teks dengan melihat urutan tertentu melalui pencocokan pola [17].
2. **Pelacakan Topik (Topic Tracking)**, Penetuan Dokumen lain yang menarik seorang pengguna berdasarkan profil dan Dokumen yang dilihat pengguna tersebut [17].
3. **Perangkuman (Summarization)**, Pembuatan rangkuman Dokumen untuk mengefisiensikan proses membaca [17].

4. **Kategorisasi (Categorization)**, Penentuan tema utama suatu teks dan pengelompokan teks berdasarkan tema tersebut ke dalam kategori yang telah ditentukan [17].
5. **Penggususan (Clustering)**, Pengelompokan Dokumen yang serupa tanpa penentuan kategori sebelumnya [17].
6. **Penautan Konsep (Concept Linking)**, Penautan Dokumen terkait dengan identifikasi konsep yang dimiliki bersama sehingga membantu pengguna untuk menemukan informasi yang mungkin tidak akan ditemukan dengan hanya menggunakan metode pencarian tradisional [17].
7. **Penjawaban Pertanyaan (Question Answering)**, Pemberian jawaban terbaik terhadap suatu pertanyaan dengan pencocokan pola berdasarkan pengetahuan [17].

1

Teks mining merupakan teknik yang digunakan untuk menangani klasifikasi, mengekstrak informasi dan pencarian informasi. Konsep teks mining yang digunakan dalam klasifikasi yaitu dokumen tekstual dengan tujuan untuk mengklasifikasi document yang sesuai dengan topic pembahasan. Perbedaan antara data mining dan text mining terletak pada *preprocessing*, pada data mining *preprocessing* berfokus pada penomoran (indexing) dan normalisasi data, sedangkan text mining berfokus pada identifikasi dan ekstraksi fitur.



2.2.3 Sentimen Analisis

1

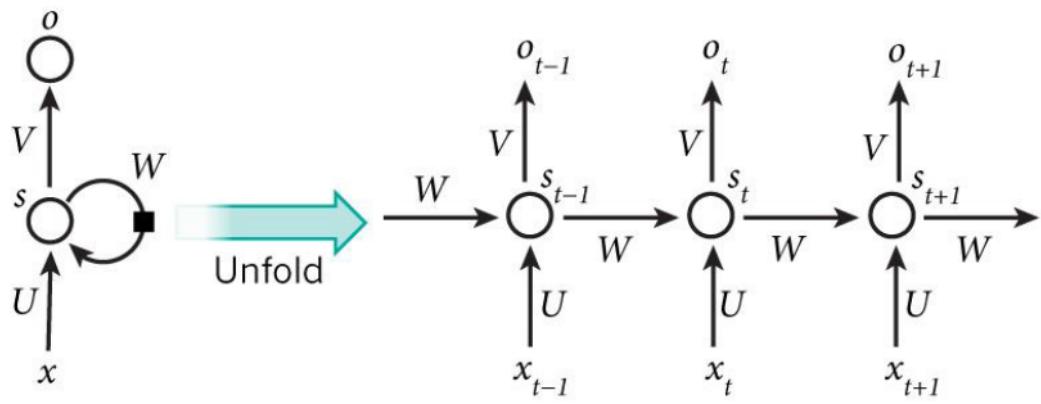
Sentiment Analysis adalah sentimen dari teks subjektif tersebut menganalisis, memproses, meringkas dan proses inferensial. Sentiment analysis saat ini dibagi menjadi penggunaan klasifikasi pembelajaran mesin dan metode klasifikasi berdasarkan aturan, metode pembelajaran mesin menggunakan kata

kata emosi sebagai klasifikasi fitur, dan kamus emosi dapat digunakan untuk mewujudkan pemilihan karakteristik sentiment dengan cepat dan efisien. Dalam kombinasi dengan pengklasifikasian dalam menyelesaikan tugas metode yang umumnya digunakan adalah Naïve Bayes (NB), Support Vector Machine (SVM) dan Entropi Maksimum (EM)[14].

2.2.4 Recurrent Neural Network

Recurrent Neural Network (RNN) merupakan bagian dari Neural Network yang di khususkan untuk memproses data yang bersambung. Untuk menyimpan informasi RNN melakukan perulangan di dalam arsitekturnya, yang membuat informasi dari masa lalu tetap tersimpan selain itu RNN juga adalah salah satu bagian dari Neural Network yang populer dalam Natural Language Processing (NLP) karena strukturnya cocok untuk pemrosesan teks.

Dari informasi di atas menunjukan bahwa RNN memiliki arsitektur yang didedikasikan untuk data berbentuk sequence dan list. Sehingga pada beberapa tahun terakhir, RNN sudah cukup terbukti dalam menyelesaikan permasalahan seperti speech recognition, machine translation, sentiment analysis, image captioning dan masih banyak lagi[15].



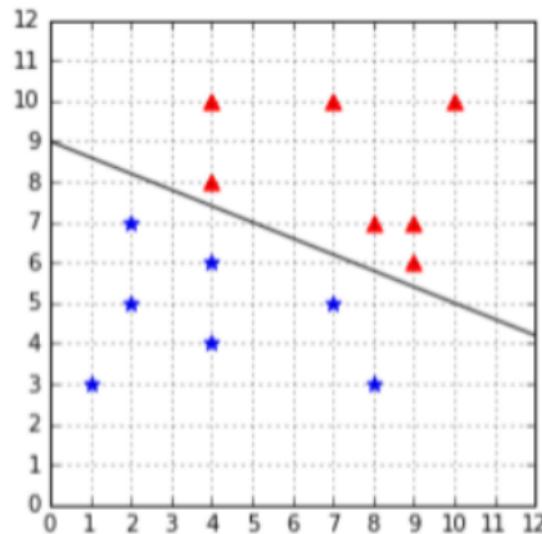
Gambar 2.2.3.1 Arsitektur RNN

2.2.5 Support Vector Machine

Support vector machine adalah sistem pembelajaran yang membuat asumsi tentang fitur berdimensi tinggi menggunakan bentuk fungsional linier dan dilatih menggunakan algoritma pembelajaran berdasarkan teori optimasi. SVM pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian konsep lanjutan di bidang pengenalan pola teknik, dan keakuratan model yang dihasilkan oleh proses transformasi dalam SVM sangat bergantung pada fungsi kernel dan parameter yang digunakan. Menurut karakteristiknya, metode SVM dibagi menjadi linier dan nonlinier. SVM linier adalah data yang dipisahkan secara linier, yaitu dua kelas dipisahkan pada hyperplane dengan margin lunak. Sementara itu, nonlinier adalah fungsi dari trik kernel di ruang dimensi tinggi.

SVM sangat cepat dan efektif untuk masalah klasifikasi teks. Dalam istilah geometris, klasifikasi biner dapat dianggap sebagai hyperplane di ruang fitur yang memisahkan titik-titik yang mewakili contoh positif dari kategori yang mewakili keadaan negatif. Klasifikasi ini dipilih selama pelatihan sebagai hyperplane unik yang memisahkan contoh positif dan negatif yang diketahui. Klasifikasi SVM memiliki keunggulan penting dalam pendekatan teoritis rasional untuk masalah tersebut[14].

SVM menggunakan 2 titik dan dua titik itu selanjutnya akan membentuk pembatas garis. Pembatas yang telah terbentuk dari 2 titik vektor ini disebut dengan hyperplane.



1 Gambar 2.2.4.1 *Hyperplane* memisahkan dua kelas.

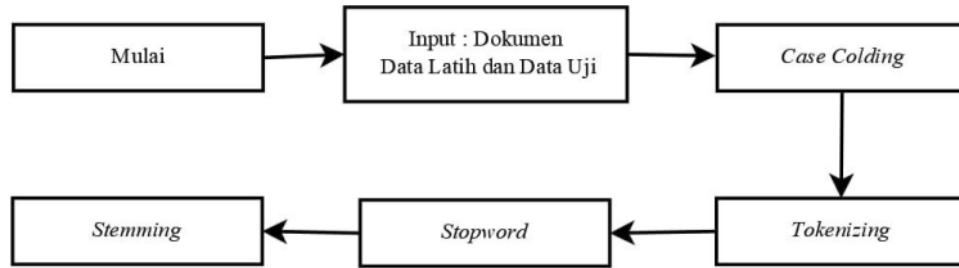
Dari dua titik vektor yang menjadi patokan hyperplane disebut dengan *support vector*. Bisa dilihat disini telah memiliki dua kelompok yang bisa disebut dengan klasifikasi, dan kemudian SVM mempunyai tugas untuk membagi dua kelompok ini sebaik mungkin dan menentukan *hyperplane* yang paling bagus.

2 **2.2.6 Preprocessing**

Text preprocessing merupakan tahap awal dimana data teks dengan noise akan dikurangi agar dapat diolah lebih lanjut, untuk mengurangi noise tersebut. Tahap text preprocessing ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi knowledge discovery sistem text mining. Data tekstual akan diproses dalam beberapa tahapan text preprocessing yaitu case

folding, tokenisasi, filtering dan stemming. Dengan dilakukannya text preprocessing akan terbentuk dataset bersih, dataset yang terbentuk dari proses ini akan memudahkan dalam pemrosesan sistem.[19]

1. **Case Folding**, *Case folding* adalah proses pada text *preprocessing* yg dilakukan dengan tujuan untuk menyeragamkan karakter dalam data. Proses *case folding* merupakan proses menjadikan semua alfabet sebagai alfabet kecil. Pada proses ini karakter-karakter 'A'-'Z' yg masih ada dalam data diubah kedalam karakter 'a'-'z'.
²
2. **Tokenisasi**, *tokenisasi* adalah proses pemecahan kata dan menghilangkan delimiter pada suatu teks. Proses ini dilakukan agar menghasilkan kumpulan kata yang berdiri sendiri.
²
3. **Filtering**, proses *filtering* adalah proses setelah tokenisasi, tujuan dari proses ini adalah untuk mengambil kata yang penting dari hasil tokenisasi yang dilakukan. Proses ini juga menjadi penentuan apakah kata akan digunakan atau tidak digunakan.
²
4. **Stemming**, *stemming* adalah proses untuk merubah kata menjadi bentuk dasar. Proses ini dilakukan dengan tujuan menyamakan bentuk data, dan dari proses ini imbuhan - imbuhan berupa prefiks, konfiks, dan sufiks akan di hilangkan.



Gambar 2.2.6.1 tahapan *pre-processing*

2.2.7 Confusion Matrix

Confusion matrix adalah *matrix* 2×2 yang merepresentasikan hasil klasifikasi bi pada suatu *dataset*. Terdapat beberapa rumus umum yang dapat digunakan untuk menghitung performa klasifikasi. Hasil dari nilai *accuracy*, *precision* dan *recall* bisa ditampilkan dalam persentase [20].

1. *Accuracy*, yang dimaksud dengan *Accuracy* adalah jumlah proporsi prediksi yang benar. Rumus perhitungan *akurasi* adalah sebagai berikut [20].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precision*, yang dimaksud dengan *precision* adalah proporsi jumlah dokumen teks relevan yang dapat di tangani pada semua dokumen yang dipilih oleh sistem. Rumus dari *precision* adalah sebagai berikut [20].

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. *Sensitivity*, yang dimaksud dengan *Sensitivity* adalah berapa % dari urutan positif yang diberi label pengklasifikasian sebagai positif [21]. Rumus dari *Sensitivity* adalah sebagai berikut.

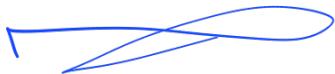
$$\text{Sensitivity} = \text{True Positive Rate (TPR)} = \frac{TP}{TP + FN}$$

4. *Specificity*, yang dimaksud dengan *Specificity* yaitu dihitung sebagai jumlah prediksi negatif yang benar dibagi dengan jumlah total negatif [21]. Rumus dari *Specificity* adalah sebagai berikut

$$\text{Specificity} = \text{True Negative Rate (TNR)} = \frac{TN}{TN + FP}$$

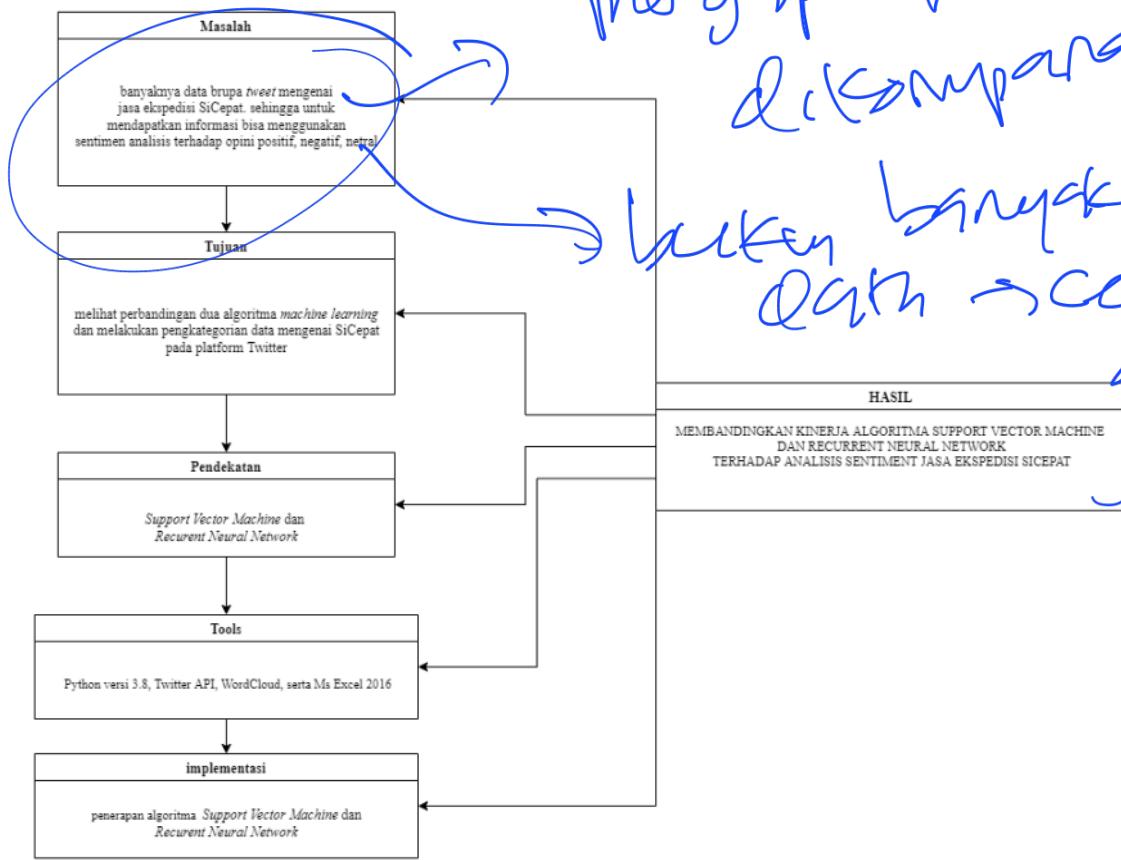
5. Yang dimaksud dengan *Balanced Accuracy* yaitu rata-rata aritmatika sensitivitas dan spesifisitas, penggunaan *Balanced Accuracy* yaitu ketika saat dihadapkan pada data yang tidak seimbang, yaitu pada saat salah satu kelas target muncul lebih banyak dari yang lainnya [22]. Rumus dari *Balanced Accuracy* adalah sebagai berikut.

$$\text{Balanced Accuracy} = \frac{\text{TPR} + \text{TNR}}{2}$$



2.3 Kerangka Konseptual

Kerangka konsep dari penelitian yang dilakukan adalah melihat tingkat akurasi yang dihasilkan algoritma *support vector machine* dan *recurrent neural network* dalam memproses sentimen dari *tweet* masyarakat yang beredar di twitter, dan melihat sentimen positif, negatif, dan netral masyarakat terhadap jasa ekspedisi SiCepat. Dalam penelitian ini akan dipaparkan tingkat akurasi dari kedua algoritma yang dipakai. Adapun kerangka konseptual yang telah digambarkan seperti pada gambar 2.3.1



Gambar 2.3.1 Kerangka Konseptual

2.4 Hipotesis Penelitian

H1 : Masyarakat Indonesia memiliki opini atau sentimen yang positif terhadap jasa ekspedisi SiCepat.

H2 : Akurasi yang dihasilkan oleh RNN lebih optimal dibandingkan akurasi yang dihasilkan oleh SVM.

H3 : Masyarakat di Indonesia lebih memilih jasa ekspedisi SiCepat dibandingkan jasa ekspedisi lainnya.

BAB III METODE PENELITIAN

Pada penelitian kali ini, peneliti akan berfokus untuk menganalisa sentimen publik terhadap jasa ekspedisi SiCepat lewat cuitan pada *platform* Twitter, lalu melakukan perbandingan terhadap algoritma Support Vector Machine dan Recurrent Neural Network. Sentimen analisis adalah salah satu bidang dalam penelitian *text mining*. Yang mana *text mining* sendiri terbagi menjadi dua yaitu pengkategorian (*classification*) dan pengelompokan (*clustering*) data. *Classification* berfungsi untuk mengelompokan data berdasarkan karakteristik yang sama, sedangkan *clustering* dipakai untuk memberikan pelabelan pada suatu kelas yang belum diketahui. Dan penelitian kali ini menggunakan *classification* atau pengkategorian, karena tujuan utama dari peneliti adalah untuk mengkategorikan *tweet* berdasarkan apakah kalimat tersebut positif, negatif, atau netral.

3

3.1 Objek dan Subjek Penelitian

Pada **penelitian** yang dilakukan kali **ini**, yang menjadi objek penelitian adalah opini dari kalangan masyarakat Indonesia terhadap jasa ekspedisi SiCepat yang dilihat dari *tweet* pada media sosial Twitter. Lalu yang menjadi subjek penelitian adalah masyarakat Indonesia pengguna Twitter yang merupakan konsumen SiCepat.

3

3.2 Populasi dan Sampel Penelitian

Populasi pada **penelitian** kali **ini** adalah semua *tweet* atau cuitan yang mengandung kata “SiCepat” pada platform media sosial Twitter. Lalu sampel yang diambil dan akan digunakan sebanyak 1.000 tweet yang mengandung kata “SiCepat”, mulai dari tahun 2020 - 2022.

3.3 Metode Pengumpulan Data

Pada penelitian kali ini, data akan dikumpulkan dengan mengekstrak *tweet* atau cuitan dari Twitter dengan menggunakan Twitter API serta bantuan Phyton. Namun sebelumnya kita harus membuat *developer account* terlebih dahulu pada website dev.twitter.com agar bisa mendapatkan akses ke Twitter API. Dengan bantuan *code* dari Phyton maka kita dapat mengekstrak *tweet* yang diperlukan, dalam hal ini *tweet* yang mengandung kata “SiCepat” secara otomatis.

3.4 Jenis dan Sumber Data

Penelitian kali ini menggunakan jenis data primer, yang mana data ini diperoleh dengan memanfaatkan Twitter API. Data yang diambil merupakan data kualitatif atau data yang berbentuk kalimat. Data-data tersebut berasal dari *tweet* yang mengandung kata “SiCepat” mulai pada tahun 2020 - 2022.

3.5 Variabel Penelitian

- a). Teks, dalam hal ini yaitu tweet dari pengguna Twitter yang mengandung kata SiCepat.
- b). Data berlabel, yaitu score yang diberikan untuk tiap teks berdasarkan kamus *lexicon*, baik itu positif, netral, maupun negatif.
- c). Klasifikasi, *tweets* atau teks yang sudah mempunyai kategori sentimen positif, netral, dan negatif.

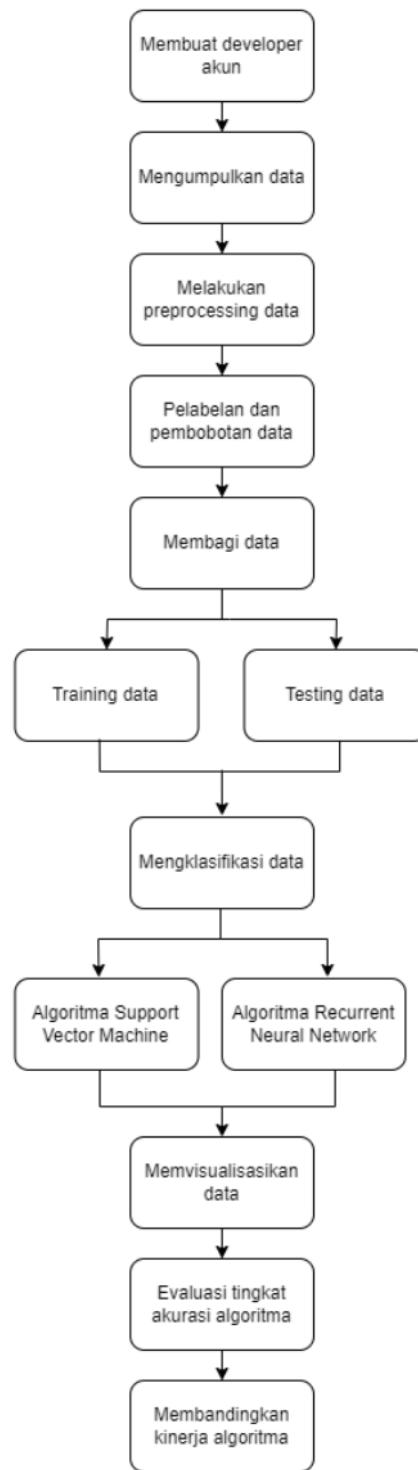
3.6 Metode Analisis Data

Pada penelitian kali ini, peneliti akan menggunakan beberapa *tools* yaitu Python versi 3.8, Twitter API, WordCloud, serta Ms Excel 2016. Adapun metode untuk menganalisa data yang digunakan pada penelitian kali ini yaitu:

1. Analisis sentimen, yang akan dipakai untuk melabelkan data ke dalam beberapa kelas, seperti sentimen positif, sentimen netral, dan sentimen negatif.
2. Metode Support Vector Machine dan Recurrent Neural Network untuk mengklasifikasikan data atau *tweet* yang bersifat positif, netral, maupun negatif.
3. Ms Excel dipakai sebagai media untuk menampung data yang akan diambil dari Twitter API menggunakan bantuan Python.
4. WordCloud dipakai untuk memvisualisasikan data yang dalam hal ini adalah kata-kata yang paling banyak atau paling sering muncul, baik yang memiliki sentimen positif, netral, maupun negatif.

3.7 Tahapan Penelitian

Penelitian kali ini akan melalui 9 tahapan untuk dapat mencapai tujuan penelitian. Di bawah adalah *flowchart* yang menggambarkan alur atau tahapan penelitian yang dilakukan oleh peneliti.



Gambar 3.7.1 *Flowchart* penelitian

1. Membuat *developer* akun Twitter

Tahapan pertama yang harus dilakukan sebelum memulaikan semua kegiatan penelitian yang lain adalah peneliti harus terlebih dahulu berhasil membuat akun *developer* Twitter serta mendapatkan persetujuan dari pihak Twitter pada website *developer* Twitter di dev.twitter.com.

2. Mengumpulkan data

Selanjutnya adalah tahapan untuk mengumpulkan data pada Twitter API dengan menggunakan bantuan Python, dan data yang diambil adalah *tweet* yang mengandung kata “SiCepat” mulai dari tahun 2020 - 2022.

3. Melakukan *preprocessing* data

Karena data yang diambil merupakan cuitan dari pengguna Twitter maka tentunya kita harus melakukan tahapan *preprocessing* data. Hal ini disebabkan karena banyak ulasan yang mengandung kata-kata yang tidak terstruktur baik itu singkatan, angka, simbol, maupun emotikon [12]. Jadi pada tahapan ini data yang sebelumnya telah diambil akan diekstrak menjadi data yang lebih terstruktur, sehingga memudahkan peneliti untuk melakukan penelitian. Tahapan *preprocessing* data yang akan dilakukan adalah sebagai berikut:

- a). Case Folding. Merupakan tahapan atau proses mengubah kata-kata ke dalam satu bentuk yang sama, contohnya kata dari huruf besar menjadi huruf kecil, maupun sebaliknya. Kali ini peneliti akan melakukan perubahan, yaitu semua kata akan diubah menjadi huruf kecil atau *lowercase* [13].
- b). Translate. Tahapan yang berfungsi untuk mengubah kata berbahasa asing menjadi bahasa Indonesia.

- c). Menghapus Simbol, Angka, serta Emotikon. Selanjutnya peneliti akan menghapus semua simbol-simbol, angka, tanda baca, dan juga *emoticon* yang ada di dalam *tweet*.
- d). Lemmatization. Proses untuk mengubah kata yang tidak lengkap atau kata yang dipersingkat menjadi kata dasar untuk lebih mudah dimengerti [12].
- e). Stopword. Proses untuk menghilangkan kata-kata yang umumnya sering dipakai serta tidak mempengaruhi suatu kalimat secara signifikan [12].
- f). Tokenizer. Tahapan untuk memisahkan kata per kata berdasarkan karakter spasi.

4. Pelabelan serta pembobotan data

Setelah berhasil melewati tahap *preprocessing* maka selanjutnya peneliti akan melakukan pelabelan serta pembobotan. Dimulai dari menghitung skor dari setiap *tweet* berdasarkan kamus *lexicon*, dan kemudian setelahnya maka *tweet* akan dilabelkan, apabila skor dari *tweet* tersebut < 0 masuk kedalam kelas negatif, apabila skor *tweet* = 0 maka dimasukan kedalam kelas netral, dan jika skornya > 0 maka masuk kedalam kelas positif.

5. Membagi data menjadi *training* dan *testing* data

Data yang telah diambil kemudian akan dibagi menjadi 2 bagian yaitu *training* data dan *testing* data. Tujuan data *training* adalah untuk melatih data yang nantinya akan dipergunakan untuk membangun model. Sedangkan *testing* data adalah data yang akan dijadikan untuk menguji performa dari model yang sebelumnya sudah dilatih tadi [12].

6. Melakukan klasifikasi data dengan algoritma SVM dan RNN

Tahap selanjutnya adalah melakukan proses klasifikasi terhadap data tersebut dengan menggunakan algoritma yang sudah ditentukan, yaitu algoritma Support Vector Machine dan Recurrent Neural Network dengan menggunakan bantuan Python.

7. Memvisualisasikan data

Pada tahapan kali ini akan dilakukan visualisasi data dengan menggunakan Wordcloud, untuk melihat kata apa yang paling sering muncul dalam *tweet* mengenai SiCepat. Nantinya kata yang paling sering muncul akan memiliki visualisasi ukuran yang lebih besar dibandingkan dengan kata-kata lainnya.

8. Evaluasi algoritma dengan *confusion matrix*

Setelah semua tahapan di atas selesai, maka selanjutnya adalah tahapan untuk mengukur kinerja algoritma yang dipakai dengan menggunakan *confusion matrix* yang memiliki empat parameter yaitu TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan juga FN (*False Negative*) [18]. Kali ini peneliti akan menggunakan confusion matrix untuk membantu menghitung *sensitivity*, *specificity*, *accuracy*, *precision*, dan juga *balanced accuracy* dari masing-masing algoritma, dengan rumus sebagai berikut:

$$\text{a). } \textit{Sensitivity} = \textit{True Positive Rate (TPR)} = \frac{\textit{TP}}{\textit{TP} + \textit{FN}}$$

$$\text{b). } \textit{Specificity} = \textit{True Negative Rate (TNR)} = \frac{\textit{TN}}{\textit{TN} + \textit{FP}}$$

$$\text{c). } \textit{Accuracy} = \frac{\textit{TP} + \textit{TN}}{\textit{TP} + \textit{TN} + \textit{FP} + \textit{FN}}$$

$$\text{d). } \textit{Precision} = \frac{\textit{TP}}{\textit{TP} + \textit{FP}}$$

$$\text{e). } \textit{Balanced Accuracy} = \frac{\textit{TPR} + \textit{TNR}}{2}$$

9. Membandingkan kinerja algoritma

Setelah selesai mengukur kinerja dari algoritma Support Vector Machine dan juga Recurrent Neural Network dengan menggunakan bantuan *confusion matrix*, maka selanjutnya kita akan membandingkan tingkat akurasi yang dihasilkan oleh masing-masing algoritma tersebut dengan mengacu kepada *score balanced accuracy*-nya. Karena *score* yang dihasilkan oleh *balanced accuracy* merupakan hasil perpaduan dari *sensitivity* dan juga *specificity*, jadi rasio yang dihasilkan lebih seimbang.

DAFTAR PUSTAKA

(takluk,
berkay!)

- [1] S. N. Salsabila, "ANALISIS FAKTOR PENENTU KEPUTUSAN KONSUMEN TERHADAP JASA PENGIRIMAN BARANG," 2021.
- [2] C. P. A. Rabbi, "Industri Logistik Tumbuh 40% Selama Pandemi Covid-19", KataData, Indonesia. Accessed: April. 21, 2022. [Online]. Available: <https://katadata.co.id/safrezifitra/berita/60e6bc288f45c/industri-logistik-tumbuh-40-selama-pandemi-covid-19>
- [3] S. N. Hakim, "ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES CLASSIFIER (NBC)," 2021.
- [4] A. V. Sudiantoro and E. Zuliarso, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier," Pros. SINTAK 2018, 2018
- [5] B. U. Manalu, "Analisis Sentimen pada Twitter Menggunakan Text Mining Skripsi Boy Utomo Manalu," Teknol. Inf. Fak. Ilk. UNSUT, 2014.
- [6] Hena, Triana (2021) ANALISIS SENTIMEN PENCITRAAN PERGURUAN TINGGI DI YOGYAKARTA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER. Skripsi thesis, STMIK AKAKOM YOGYAKARTA.
- [7] Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computational Science*, 27, 386–393. <https://doi.org/10.1016/J.JOCS.2017.11.006>
- [8] Muludi, K., Akbar, M. S., Shofiana, D. A., & Syarif, A. (2021). Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(4), 339. <https://doi.org/10.22146/ijccs.66016>

- [9] Merinda Lestandy, Abdurrahim Abdurrahim, & Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802–808. <https://doi.org/10.29207/resti.v5i4.3308>
- [10] Sharazita Dyah Anggita, & Ikmah. (2020). Algorithm Comparation of Naive Bayes and Support Vector Machine based on Particle Swarm Optimization in Sentiment Analysis of Freight Forwarding Services. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(2), 362–369. <https://doi.org/10.29207/resti.v4i2.1840>
- [11] Welcome to UPN Veteran Jatim Repository - UPN Veteran Jatim Repository. (n.d.). Retrieved April 25, 2022, from <http://repository.upnjatim.ac.id/4727/>
- [12] A. Faadilah, “ANALISIS SENTIMEN PADA ULASAN APLIKASI TOKOPEDIA DI GOOGLE PLAY STORE MENGGUNAKAN METODE LONG SHORT TERM MEMORY,” 2020.
- [13] S. N. Kane, A. Mishra, and A. K. Dutta, “Pre-processing Tasks in Indonesian Twitter Messages,” *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2016
- [14] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). ¹ SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 31–37. <https://doi.org/10.33365/JDMSI.V2I1.1021>
- [15] Fатуровман, F., Иrawan, B., & Setianingsih, C. (2020). Analisis Sentimen Pada Bpjs Kesehatan Menggunakan Recurrent Neural Network. *E-Proceeding of Engineering*, 7(2), 4545–4552.
- [16] Merinda Lestandy, Abdurrahim Abdurrahim, & Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802–808. <https://doi.org/10.29207/resti.v5i4.3308>
- [17] ⁴ Apa itu Text Mining? – School of Information Systems. (n.d.). Retrieved May 1, 2022, from <https://sis.binus.ac.id/2021/04/23/apa-itu-text-mining/>
- [18] A. Patel and A. K. Tiwari, “Sentiment Analysis by using Recurrent Neural Network,” 2019.

- [19] Ardiani, L., Sujaini, H., Hadari Nawawi, J. H., & Barat, K. (2020). *Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak*. JUSTIN (*Jurnal Sistem Dan Teknologi Informasi*), 8(2), 183–190. <https://doi.org/10.26418/JUSTIN.V8I2.36776>²
- [20] Andika, L. A., Azizah, P. A. N., & Respatiwulan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, 2(1), 34. <https://doi.org/10.13057/ijas.v2i1.29998>³
- [21] *Basic evaluation measures from the confusion matrix – Classifier evaluation with imbalanced datasets.* (n.d.). Retrieved May 8, 2022, from <https://classeval.wordpress.com/introduction/basic-evaluation-measures/>
- [22] *Balanced Accuracy: When Should You Use It? - neptune.ai.* (n.d.). Retrieved May 8, 2022, from <https://neptune.ai/blog/balanced-accuracy>

SIS_SENTIMENT_JASA_EKSPEDISI_SICEPAT_- _Ferren_Hillary_Kalalo.pdf

ORIGINALITY REPORT



PRIMARY SOURCES

1	ejurnal.teknokrat.ac.id Internet Source	7%
2	jurnal.untan.ac.id Internet Source	4%
3	Dspace.Uii.Ac.Id Internet Source	3%
4	sis.binus.ac.id Internet Source	3%

Exclude quotes Off

Exclude bibliography Off

Exclude matches < 3%