

KOMPARASI ALGORITMA KLASIFIKASI UNTUK SENTIMEN ANALYSIS DJ.ID PADA TWITTER

Medy Wisnu Prihatmono¹⁾, Ida²⁾

1. Sistem Informasi, STMIK Profesional Makassar
email: medywisnuprihatmonoprofesional@gmail.com
2. Sistem Informasi, STMIK Profesional Makassar
email: idadamulyadi7973@gmail.com

Abstrak

Saat ini pendapat masyarakat menjadi sumber yang penting dalam pengambilan keputusan akan suatu produk atau jasa. Sentimen analisis berfungsi sebagai informasi hasil dari mengekspresikan secara tekstual terhadap pendapat sebuah masalah atau untuk mengidentifikasi kecenderungan hal yang terjadi di pasar. Tujuan dari penelitian ini adalah mengetahui seberapa besar sentimen analisis yang positif, negatif atau netral yang dilakukan oleh masyarakat yang bertransaksi online pada situs di website JD.id, berdasarkan komentar yang ada di sosialmedia twitter. Alasan utama menggunakan twitter karena komentar hanya dibatasi sebanyak 140 karakter sehingga mudah untuk dikelola. Metode klasifikasi yang digunakan dalam penelitian ini adalah SVM dan logistic regression yang dimana hasil yang didapatkan Support Vector Machine untuk AUC adalah 0.587, sedangkan dengan menggunakan metode neural network hasil yang diperoleh untuk AUC adalah 0.716.

Kata Kunci : SVM, Neural Network, Sentimen Analisis

A. PENDAHULUAN

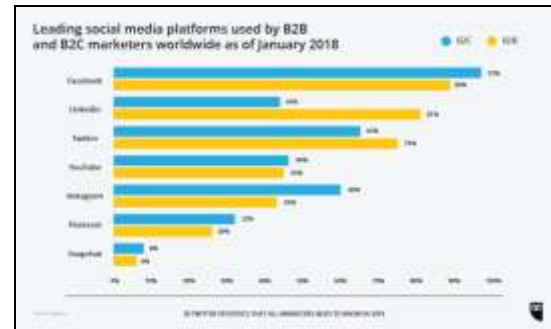
Perkembangan Internet hingga ditahun 2019 sangat pesat sekali, dengan ditandai semakin mudahnya mengakses internet dimana saja. Mulai dari kantor, kampus, sekolah, dan semua tempat umum biasanya sudah dilengkapi dengan banyaknya akses wireless Internet dengan kecepatan yang bervariasi. Perkembangan teknologi smartphone, dan teknologi jaringan 4G membuat masyarakat lebih mudah dan cepat mengakses informasi melalui internet. Kemudahan akses internet ini mengubah kebiasaan

masyarakat yang biasanya mengakses informasi melalui media cetak menjadi melalui internet. Sementara dengan berkembang pesatnya pengguna internet, semakin berkembang pula media sosial. Media sosial merupakan media online, dimana penggunanya bisa mengakses informasi, berpartisipasi, berbagi dan menciptakan isi. Salah satu jejaring sosial yang marak digunakan adalah twitter. Melalui twitter pengguna bisa mengekspresikan tulisannya dalam 140 karakter. Saat ini twitter telah mencapai sekitar 400 juta pengguna aktif pada tahun 2014. Di Indonesia sendiri, pengguna

twitter hingga tahun 2014 mencapai 20 juta pengguna aktif [1]. Ini membuat jejaring sosial tersebut menjadi channel penting untuk berkampanye. Didalam perkembangannya twitter itu memiliki banyak keunggulan dibandingkan dengan media sosial lainnya, banyak konten-konten yang bisa dimanfaatkan perusahaan / perorangan untuk memaksimalkan twitter sebagai media promosi seperti following, tweeting, replies, retweet dan trending topics. [2]. Dengan memanfaatkan semua fitur dari twitter, bagi perusahaan dapat melihat perkembangan evaluasi produk dan untuk melihat reaksi pasar, bisa menjadi sarana edukasi konsumen untuk product review, serta mengkomunikasikan kelebihan atau kekurangan dibandingkan dengan pesaing lain. Dengan penggunaan sosial media khususnya twitter, yang dipakai pada strategi marketing model bisnis e-commerce untuk B2B dan B2C. Pada bagian B2B disini fungsinya transaksi yang dilakukan secara elektronik maupun fisik dan terjadi antara entitas, misalnya ketika dua perusahaan mengadakan transaksi jual beli secara online, dengan pembayaran yang tersedia menggunakan kartu kredit.

Sedangkan pada bagian B2C fungsinya bisnis yang dilakukan antara

pelaku bisnis dengan konsumen, misalnya produsen menjual produk ke konsumen secara online. Pada gambar 1. dibawah ini ditampilkan sejumlah sosial media yang beraktivitas pada area B2B dan B2C



Gambar 1 Statistik Twitter untuk bisnis.

Dalam penelitian ini penulis membahas tentang JD.Id. JD.ID adalah perusahaan mal daring yang beroperasi di Jakarta, Indonesia. JD.ID dibentuk sebagai kerja sama antara JD.com dan sebuah partner Indonesia. JD.com adalah salah satu toko B2C daring terbesar di China secara jumlah transaksi. JD.com adalah “rival terbesar Alibaba, sang pemimpin pasar China. Pada penelitian sebelumnya yang menggunakan data twitter sudah banyak dilakukan, berikut ini beberapa review jurnal tentang penggunaan data twitter.

Penelitian yang dilakukan oleh [3] yang dimana membuat Aplikasi Deteksi Kemacetan yang dirancang menggunakan beberapa diagram UML. Aplikasi Deteksi Kemacetan ini terkoneksi dengan Hadoop, Flume, Hive, Derby, dan Twitter API untuk dapat melakukan streaming data

twitter secara real-time dalam pengumpulan datanya. Aplikasi ini dibuat dengan bahasa pemrograman Java, lengkap dengan fitur untuk melihat isi data, melakukan pencarian data tabel tweet berdasarkan ID, melakukan analisa keadaan lalu lintas suatu daerah, dan melakukan analisa sentimen terhadap suatu tweet pada tabel, dan menampilkan hasil analisa berdasarkan data pada tabel. Hasil penelitian ini dilakukan pengujian data dengan membandingkan data Aplikasi Deteksi Kemacetan dengan data Google Maps.

Kemudian penelitian yang dilakukan oleh [4] Pada penelitian ini diusulkan metode baru untuk mendeteksi antara bot spammer dan legitimate user dengan mengintegrasikan sentimen analysis berdasarkan emosi dan time interval entropy. Pendekatan gabungan knowledge-based dan machine learning-based digunakan untuk mengklasifikasi tweet yang memiliki sentimen positif, negatif dan tweet netral. Selanjutnya kumpulan timestamp digunakan untuk menghitung time interval entropy dari *tiap akun*. Hasil percobaan menunjukkan bahwa *precision* dan *recall* dari metode yang diusulkan mencapai 83% dan 91%.

Kemudian penelitian yang dilakukan oleh [5] yang dimana dokumen teks yang

digunakan dalam penelitian berasal Twitter tentang opini masyarakat mengenai penyedia layanan telekomunikasi seluler. Metode yang digunakan adalah Support Vector Machine dengan menggunakan Lexicon Based Features sebagai pembaharuan fiturnya selain memakai fitur TF-IDF. Data yang digunakan pada penelitian ini sebanyak 300 data yang dibagi menjadi dua jenis data dengan perbandingan 70% untuk data latih dan 30% untuk data uji. Hasil akurasi sistem yang diperoleh dari analisis sentimen dengan metode Support Vector Machine dan Lexicon Based Features sebesar 79% menggunakan nilai degree sebesar 2, nilai konstanta learning rate 0,0001, serta jumlah iterasi maksimum sebanyak 50 kali. Kemudian

Penelitian yang dilakukan oleh [6] Makalah ini berkontribusi pada analisis sentimen bidang penelitian. Pertama, langkah-langkah praproses yang diperlukan untuk mengekstrak fitur dari data Twitter dijelaskan. Kemudian, berbasis topik metode diusulkan untuk memperkirakan sentimen tweet. Metode ini mengharuskan untuk mengekstrak topik dari set data pelatihan, dan melatih model untuk masing-masing topik ini. Metode ini memungkinkan untuk meningkatkan akurasi estimasi sentimen dibandingkan dengan

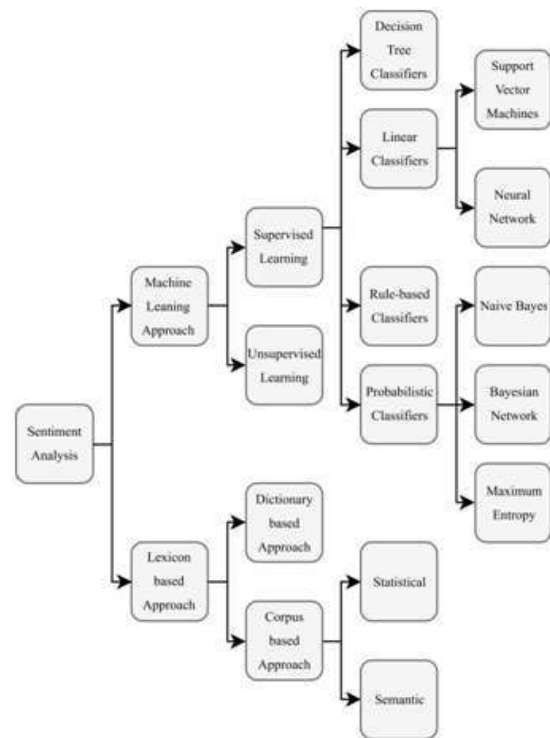
menggunakan model tunggal untuk setiap topik..

Berdasarkan dari beberapa referensi diatas, maka penulis mencoba membuat penelitian yang berjudul "komparasi algoritma klasifikasi untuk sentimen analysis dj.id pada twitter" dengan menggunakan data twitter dengan periode tertentu kemudian diproses dengan menggunakan algoritma support vector machines dan Neural Network dengan menggunakan tools berbasis open source.

B. METODE PENELITIAN

Analisis Sentimen

Analisa sentimen dapat diartikan sebagai klasifikasi sentimen dari data atau teks yang kebanyakan tidak terstruktur menggunakan kombinasi dari NLP (Natural Language Processing) dan teknik komputasional [7]. Sentimen Analysis pada dasarnya dibagi menjadi dua bagian yaitu: Machine learning approach dan lexicon based approach, yang dimana kedua bagian pokok dasar itu, mempunyai sub komponen lagi *didalamnya*, untuk lebih jelasnya bisa dilihat pada gambar 2. berikut ini [8] :



Gambar 2 Klasifikasi Sentimen Analisis

Text Mining

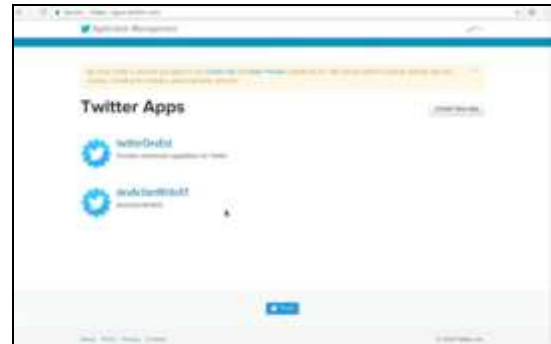
Text Mining juga dikenal sebagai penemuan pengetahuan dari database berbasis konsep tekstual [9]. Text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam data mining. Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan dalam text mining adalah sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur.

Adapun tugas khusus dari text mining antara lain yaitu pengkategorisasian teks dan pengelompokkan teks [10].

Twitter API

Twitter adalah situs jejaring sosial yang memungkinkan setiap para user/pengguna berbagai terhadap sebuah informasi dalam bentuk teks sejumlah 140 karakter [Phuvipadawat and Murata, 2010]. Twitter memiliki Application Programming Interface (API) yang diperuntukkan developer dapat mengembangkan aplikasi sesuai dengan kebutuhannya masing - masing [11]. Twitter REST API: Terdiri dari Twitter REST dan Twiter Search. Twitter REST memberikan core data dan core twitter objects. Twitter search berfungsi untuk melakukan pencarian mengenai suatu instance objek Twitter maupun mencari trend. Twitter Streaming API: API ini biasa digunakan untuk penggalian data karena melalui API ini informasi bisa didapatkan secara realtime dengan volume yang sangat tinggi. Pada penelitian ini sendiri, Twiter Streaming API akan digunakan untuk mengumpulkan data penelitian. Untuk mendapatkan Twitter API tentunya, langkah awalnya mempunyai akun di twitter terlebih dahulu, apabila belum memiliki akun di twitter maka kunjungi di web situs

<http://www.twitter.com> untuk mendaftar atas kepemilikan sebuah akun twitter. Untuk membuat Twitter API kunjungi situs <https://apps.twitter.com> untuk melakukan registrasi, pada gambar.3 dibawah ini, merupakan halaman utama untuk mendaftarkan twitter API.



Gambar 3 Halaman daftar Twitter API

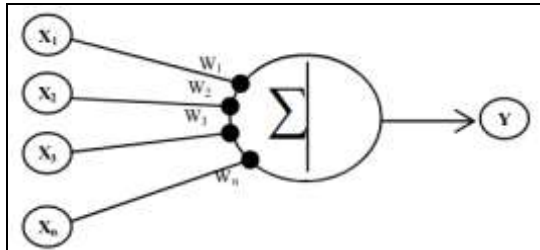
Preprocessing

Preprocessing Text merupakan suatu permulaan awal dalam melakukan sebuah text mining, yang dimana tujuan utamanya adalah merubah data yang sebelumnya tidak terstruktur menjadi terstruktur (Perdana). Didalam Preprocessing ada beberapa tahap didalamnya yaitu : casefolding, tokenizing, filtering, dan stemming.[12]

Neural Network

Metode neural network digambarkan seperti studi proses di otak manusia, yang dimana jaringan-jaringan tersebut memiliki pengetahuan-pengetahuan melalui dari proses pembelajaran. Pada gambar

no.4 digambarkan bagian untuk koneksi antara neuron yang di maksud sebagai bobot sinaptik di gunakan untuk mengingat kemampuan [Kusumadewi 2004]:



Gambar 4. Komunikasi Antar neuron

Pada Gambar xx diperlihatkan bahwa neural network terdiri dari beberapa satuan pemrosesan dari neuron. Dimana Y sebagai output yang menerima inputan dari neuron $x_1, x_2, x_3, \dots, x_n$ dengan bobot $w_1, w_2, w_3, \dots, w_n$. Hasil penjumlahan seluruh impuls neuron dibandingkan dengan nilai ambang tertentu melalui fungsi aktivasi f setiap neuron. Fungsi aktivasi digunakan sebagai penentu keluaran suatu neuron.

Support Vector Machines

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Support Vector Machine masuk kelas supervised learning, dimana dalam implementasinya perlu adanya tahap pelatihan

menggunakan sequential training SVM dan disusul tahap pengujian [13]

Pada dasarnya konsep dasar dari klasifikasi dari metode SVM adalah mencari hyperplane terbaik sebagai pemisah kedua kelas data. Persamaan dari Support Vector Machine adalah [5] :

$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b$$

w : parameter hyperplane yg dicari

x : titik data masukan SVM

a_i : nilai bobot setiap titik data

$K(x, x_i)$: fungsi kernel

b : parameter hyperplane yg dicari

C. HASIL DAN PEMBAHASAN

Pada penelitian ini, data yang diambil yaitu dengan memanfaatkan dari akun twitter yang dimiliki dari pihak penulis itu sendiri, yang dimana juga memanfaatkan API twitter dengan terlebih dahulu memiliki token akun. Kemudian data diambil dengan Pada penelitian ini, data yang diambil yaitu dengan memanfaatkan dari akun twitter yang dimiliki dari pihak penulis itu sendiri, yang dimana juga memanfaatkan API twitter dengan terlebih dahulu memiliki token akun. Kemudian data diambil dengan menggunakan idkunci yaitu "jdid", kemudian data dari hasil pencarian tersebut kemudian di proses dengan menggunakan dua

```

graph TD
    subgraph Data_Preparation [Data Preparation]
        Twitter[Icon] --> Auth1[Autentifikasi berdasarkan Token Akun]
        Auth1 --> Data[(Twitter Data)]
    end
    subgraph Input
        Data --> DataTeks[/Data Teks/]
    end
    subgraph Proses [Proses]
        DataTeks --> Auth2[Autentifikasi berdasarkan Token Akun]
        Auth2 --> ML{Algo. Machine Learning}
        ML --> End([End])
    end
    subgraph Output
        ML --> Sentimen[Sentimen Positif & Sentimen Negatif]
        Sentimen --> Hasil[Hasil Sentimen analisis dalam bentuk tabel]
    end

```

Pada bagian ini akan menjelaskan hasil - hasil yang diperoleh dengan menggunakan data dari twitter, yang dimana menampilkan hasil wordcloud, kemudian hasil analisis sentimen yang bersifat negatif dan positif, dan hasil perbandingan hasil klasifikasi dari menggunakan metode support vector machines dan neural network.

WordCloud Dengan menggunakan query word list atau kata kunci dengan judul "jd.id", dengan jumlah max tweets yang digunakan sebanyak 100 data dokumen, maka data akan di unduh akan

Output dari hasil Preprocess Text dibuat dalam bentuk WordCloud yang hasilnya ditampilkan pada gambar 6 dibawah berikut ini :

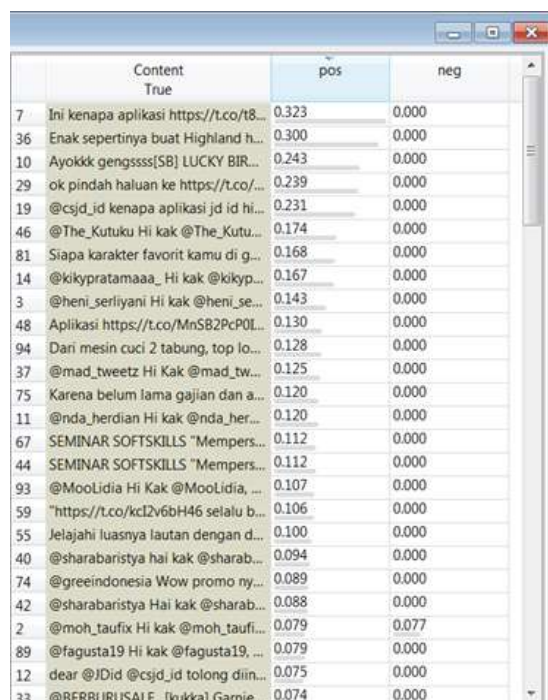


Pada gambar 6, menjelaskan tentang kata-kata yang memiliki bentuk huruf yang lebih besar memiliki arti kata & bobot menampilkan frekuensi lebih besar yang diperbincangkan, diurutkan

berdasarkan topiknya, dalam hal ini kata kunci yang dimaksud adalah "jdid".

Sentimen Analisis

Data yang dipakai sebanyak 100 data set yang diperoleh dari twitter, setelah dianalisis terhadap sentimen analisis yang di proses maka didapatkan hasil dari positif dan negatif. untuk mendapatkan nilai ini terlebih dahulu menggunakan module NLTK terdiri dari metode Liu Hu : lexicon-based sentiment analysis dan vader : lexicon -and rule-based sentimen analysis hasil yang diperoleh, dapat dilihat pada gambar 7



	Content	True	pos	neg
7	Ini kenapa aplikasi https://t.co/t8...		0.323	0.000
36	Enak sepeertinya buat Highland h...		0.300	0.000
10	Ayokkk gengssss[SB] LUCKY BIR...		0.243	0.000
29	ok pindah haluan ke https://t.co/...		0.239	0.000
19	@csjd_id kenapa aplikasi jd id hi...		0.231	0.000
46	@The_Kutuku Hi kak @The_Kutu...		0.174	0.000
81	Siapa karakter favorit kamu di g...		0.168	0.000
14	@kikypratamaaa_ Hi kak @kikyp...		0.167	0.000
3	@heni_serliyani Hi kak @heni_se...		0.143	0.000
48	Aplikasi https://t.co/MnSB2PcPOL...		0.130	0.000
94	Dari mesin cuci 2 tabung, top lo...		0.128	0.000
37	@mad_tweetz Hi Kak @mad_tw...		0.125	0.000
75	Karena belum lama gaji dan a...		0.120	0.000
11	@nda_herdian Hi kak @nda_her...		0.120	0.000
67	SEMINAR SOFTSKILLS "Mempers...		0.112	0.000
44	SEMINAR SOFTSKILLS "Mempers...		0.112	0.000
93	@MooLidia Hi Kak @MooLidia_...		0.107	0.000
59	"https://t.co/kcl2v6bH46 selalu b...		0.106	0.000
55	Jelajahi luasnya lautan dengan d...		0.100	0.000
40	@sharabaristya hai kak @sharab...		0.094	0.000
74	@greeindonesia Wow promo ny...		0.089	0.000
42	@sharabaristya Hai kak @sharab...		0.088	0.000
2	@moh_taufix Hi kak @moh_taufi...		0.079	0.077
89	@fagusta19 Hi kak @fagusta19, ...		0.079	0.000
12	dear @JDid @csjd_id tolong diin...		0.075	0.000
33	@BERBURUSALE [kukka] Garnie...		0.074	0.000

Gambar 7 sentimen-analisis "jdid"

Confusion matrix

Pada penelitian ini, penulis menggunakan confusion matrix yang dimana fungsi utama dari confusion

matrix ini adalah untuk mencari dan mengevaluasi metode tertentu, dengan syarat dapat dilakukan setelah data training yang sudah lengkap. Confusion matrix ini akan digunakan untuk menguji data yang menggunakan metode Support Vector Machine dan Neural Network. Confusion matrix pada metode Support Vector Machine yang dimana data pada posisi baris kiri, itu menandakan jumlah index sebanyak 100 data klasifikasinya yang dipakai hasil dari pengambilan data pada sosial media twitter, dan dengan cara yang sama dalam mengambil data, pada baris kolom bagian atas merupakan hasil prediksi yang diharapkan yang ditandai dengan kotak warna biru (hasil testing data), pada contoh misalnya kasus index data berjumlah 100, pada baris ke tiga dengan item "Jd.id" jumlah prediksi yang benar dikerjakan dengan hasilnya 35.0% , pada bagian lain masih banyaknya data juga yang belum diproses dengan baik dengan metode diatas, adapun hasilnya dapat dilihat pada gambar 8 dibawah ini :



Gambar 8 confision matrix dari metode Support Vector Machine

Confusion matrix yang dipakai dengan menggunakan metode Neural Network yang dimana cara memprosesnya sama dengan diatas, yaitu mengambil data sebanyak 100 set dari situs sosial twitter, contoh hasil prediksi yang sudah diproses didapat pada data index sebelah kiri yaitu CS JD bernilai 37.7 %, Jd.id bernilai 40.8% dan Xiao mencapai bernilai 100 % , apabila diamati nilai yang dianalisis, diperoleh hasil yang lebih bagus sedikit dari pada menggunakan metode Support Vector Machine, untuk lebih jelasnya bisa dilihat pada Gambar 9 dibawah ini :



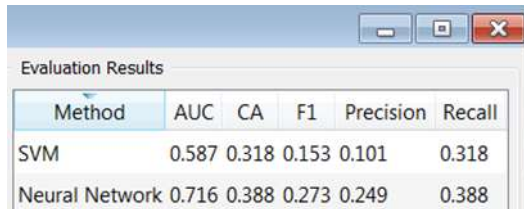
Gambar 9 confusion matrix dari metode Neural Network

Komparasi SVM dan NN

Hasil yang dicapai dengan menggunakan kedua metode Support Vector Machine dan Neural Network tersebut, interval nilai di mulai dari 0 sampai 1, pada kolom AUC merupakan bahasan tentang keakuratan dengan proses curva yang diterima dalam proses klasifikasi dari data twitter yang diperoleh, yang dimana hasil yang didapat adalah dengan menggunakan metode

SVM yaitu 0.587 sedangkan dengan metode Neural Network nilai yang didapat yaitu 0.716. Pada kolom CA yang berarti nilai yang diperoleh untuk klasifikasi secara akurat pada contoh data sampel yang dipakai, dalam hal ini dengan menggunakan metode Support Vector Machine nilai yang didapat yaitu 0.318 sedangkan pada metode Neural Network bernilai 0.388. Pada kolom F1 nilai yang diperoleh merupakan nilai harmoni dari perbandingan rata-rata antara precision and recall, yang dimana F1 untuk metode Support Vector Machine bernilai 0.153 sedangkan untuk metode Neural Network bernilai 0.273. Pada kolom dari Precision yang menceritakan jumlah sampel yang berkategori positif, yang diklasifikasikan benar sebagai sampel positif, dibagi dengan total sampel yang diklasifikasikan sebagai sampel positif, terlihat hasil yang dicapai dengan menggunakan metode Support Vector Machine bernilai 0.101, dan dengan menggunakan metode Neural Network nilai yang dicapai sebagai 0.249. Pada kolom dari recall jumlah sampel yg diklasifikasikan yang berlebih positif dibagi seberapa banyak sampel positif ada testing set, testing set adalah bagian pembelajaran dengan kelas label yaitu positif dan negatif. hasil yang diperoleh kembali dengan metode Support Vector

Machine berkisar 0.318 dan dengan metode Neural Network bernilai 0.388, untuk lebih jelasnya hasil perbandingan antara dua metode tersebut, dapat dilihat pada gambar 10 dibawah ini :



Method	AUC	CA	F1	Precision	Recall
SVM	0.587	0.318	0.153	0.101	0.318
Neural Network	0.716	0.388	0.273	0.249	0.388

Gambar 10. Hasil Perbandingan Komparasi

D. KESIMPULAN DAN SARAN

Penulis menyimpulkan penelitian sistem pengklasifikasian Tweet yang sudah dilakukan adalah sebagai berikut :

- Melakukan klasifikasi tweets dengan Metode Neural Network memiliki hasil output yang lebih baik dari pada metode Support Vector Machines dengan berdasarkan acuan AUC,CA,F1,Precision dan Recall.
- Hasil sentimen analisis terhadap produk penjualan e-commerce "JD.id" cenderung bersifat positif ketimbang negatif yang dimana nantinya bisa menjadi testimoni yang baik bagi kedua belah pihak antara penjual dan pembeli terhadap suatu produk layanan yang diperdagangkan.

Penulis menyarankan pengembangan penelitian lebih lanjut system pengklasifikasian Tweet sebagai berikut:

- Peneliti menyarankan untuk penelitian selanjutnya untuk penambahan jumlah data latih dan data uji untuk mendapatkan hasil yang lebih baik saat klasifikasi tweets.
- Peneliti menyarankan untuk penelitian selanjutnya bahasa yang digunakan juga tidak hanya bahasa Indonesia tetapi dapat menggunakan bahasa daerah atau bahasa asing seperti bahasa inggris dan bahasa lainnya.
- Peneliti menyarankan untuk penelitian selanjutnya pada tahap pemrosesan teks ditambah fitur stemming untuk mendapatkan hasil yang lebih baik
- Peneliti menyarankan untuk menggunakan metode lain dalam mengklasifikasikan hasil data tweet

REFERENSI

- Luckman, E. (2014, June 5). Technasia. Retrieved from Technasia: <http://id.technasia.com/pengguna-aktif-twitter-di-indonesia-hampir-mencapai-angka-20-juta/>.
- Rosanti Utami Dewi, "penggunaan jejaring sosial twitter sebagai media promosi pada restoran ranjang 69" Commed : Jurnal Komunikasi dan Media, Vol. 1 No.2 Februari 2017. ISSN. 2527-8673.

- [3] Busman, "Analisa Sentimen Menggunakan Data Twitter, Flume, Hive Pada Hadoop dan Java untuk Deteksi Kemacetan di Jakarta." JOIN (Jurnal Online Informatika) Volume 3 No. 1 | Juni 2018 : 1-9, DOI: 10.15575/join.v3i1.141.
- [4] Christian Sri Kusuma Aditya, "Deteksi Bot Spammer pada Twitter Berbasis Sentiment Analysis dan Time Interval Entropy". Jurnal Buana Informatika, Volume 7, Nomor 3, Juli 2016: 179-186.
- [5] Umi Rofiqoh, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 1, No. 12, Desember 2017, hlm. 1725-1732. e-ISSN: 2548-964X
- [6] Pierre FICAMOS, "A Topic based Approach for Sentiment Analysis on Twitter Data" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, 2016.
- [7] S. Kamran, M. Shaikh, A. Naseem, and P. Kamble, "Exploiting Social Media Data for Traffic Monitoring Using the Techniques of Data Mining," Int. J. Innov. Res. Comput. Commun. Eng., vol. 4, no. 3, 2016.
- [8] Arvind Singh Raghuwanshi, Polarity Classification of Twitter Data using Sentiment Analysis, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 6 ISSN: 2321-8169.
- [9] Feldman, R & Dagan, I. (1995) Knowledge discovery in textual databases (KDT). Dalam Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, AAAI Press, 112-117
- [10] Triawati, Candra; Bijaksana, M.Arif; Indrawati, Nur; Saputro, Widyanto Adi. (2009) Pemodelan Berbasis Konsep Untuk Kategorisasi Artikel Berita Berbahasa Indonesia. Dalam Seminar Nasional Aplikasi Teknologi
- [11] Ahmad Fathan Hidayatullah, Azhari SN, (2014). Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter
- [12] Agusta, L., Kristen, U. and Wacana, S., 2009. Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. Konferensi Nasional Sistem dan Informatika 2009, (KNS&I09-036), pp.196–201.
- [13] Santosa, B., 2015. Tutorial Support Vector Machine. [e-book] Surabaya: Institut Teknologi Sepuluh Nopember. Tersedia di: Google Cendekia <<https://scholar.google.com/>> [Diakses 28 Februari 2017]