

Analisis Sentimen Pada Twitter @Ovo_Id dengan Metode Support Vectore Machine (SVM)

Hendri Sulastomo¹, Ramadiansyah², Khalil Gibran³, Efrian Maryansyah⁴, Aththorihq Tegar⁵

Universitas Nusa Mandiri, Indonesia

e-mail: hendri.sulastomo@gmail.com¹, diansyahrama007@gmail.com²,

kgibran85@gmail.com³, efrianmaryansyah220488@gmail.com⁴,

dedi.eis@nusamandiri.ac.id⁵

Abstract

Social networking helps internet users communicate. This is because social network users can convey messages by utilizing the facilities prepared by each social media. Social media users' messages can be used in various ways, such as a review of a product or a review of a problem in politics or current social problems. This can be done by analyzing the sentiments of social media users. The support vectore machine method is one method that can be used to analyze sentiment. In sentiment analysis using the support vectore machine method, it is done by classifying sentiment into compliant or not compliant classes. The accuracy rate of sentiment analysis for @Ovo_ID using the support vectore machine method is 94% using 1000 tweet data.

Keywords: Analisis Sentiment, Tweet, Ovo_ID, Support Vectore Machine (SVM), Ovo

Abstrak

Jejaring sosial membantu pengguna internet dalam berkomunikasi. Hal ini dikarenakan pengguna jejaring sosial dapat menyampaikan pesan dengan memanfaatkan fasilitas yang disiapkan oleh setiap media sosial. pesan-pesan para pengguna media sosial dapat dimanfaatkan dalam berbagai hal, seperti rievew terhadap suatu produk atau rievew terhadap suatu masalah pada politik atau masalah sosial sekarang ini. Hal ini bisa dilakukan dengan menganalisis sentiment para pengguna sosial media. Metode support vectore machine merupakan salah satu metode yang dapat digunakan untuk menganalisis sentiment. Dalam sentiment analisis menggunakan metode support vectore machine dilakukan dengan cara mengklasifikasikan sentiment kedalam kelas komplain atau kelas non komplain. Tingkat akurasi analisis sentiment terhadap @Ovo_ID menggunakan metode support vectore machine adalah 94% dengan menggunakan 1000 data kicauan.

Kata kunci: Analisis Sentiment, Tweet, Support Vectore Machine (SVM), Ovo, Ovo_ID

1. PENDAHULUAN

Berdasarkan laporan *We Are Social*, jumlah pengguna aktif media sosial di Indonesia sebanyak 191 juta orang pada Januari 2022. Jumlah itu telah meningkat 12,35% dibandingkan pada tahun sebelumnya yang sebanyak 170 juta orang. Melihat trennya, jumlah pengguna media sosial di Indonesia terus meningkat setiap tahunnya. Walau demikian, pertumbuhannya mengalami fluktuasi sejak 2014-2022. Kenaikan jumlah pengguna media sosial tertinggi mencapai 34,2% pada 2017. Hanya saja, kenaikan tersebut melambat hingga sebesar 6,3% pada tahun lalu. Angkanya baru meningkat lagi pada tahun ini. Adapun, Whatsapp menjadi media sosial yang paling banyak digunakan masyarakat Indonesia. Persentasenya tercatat mencapai 88,7%. Setelahnya ada Instagram dan Facebook dengan persentase masing-masing sebesar

84,8% dan 81,3%. Sementara, proporsi pengguna TikTok dan Telegram berturut-turut sebesar 63,1% dan 62,8%.

Media Social Twitter memiliki jumlah sekitar 18.45 juta user di Indonesia dari 7.91 juta user global (Kominfo, Indonesia). Jumlah yang cukup besar tersebut menimbulkan banyak ciutan dari para penggunanya tentang berbagai hal seperti, pendidikan, hiburan, pekerjaan, dan termasuk juga politik. Salah satu issue yang menjadi *trending topic* di Twitter pada tahun 2022 adalah tentang *complain* dan *not compliant*. *Compliant* dan *Not Compliant* adalah suatu kepuasan yang konsumen sampaikan kepada Ovo_ID agar pelayanan mereka selalu memuaskan para konsument. Penulis pada penelitian ini akan melakukan analisis sentiment para pengguna Twitter terhadap kepuasan pelanggan akan pelayanan Ovo_ID. Dengan input beberapa data tweet dalam Bahasa Indonesia, akan dilakukan klasifikasi dengan algoritma *SVM (Support Vector Machine)* [1][2] untuk menentukan apakah tweet tersebut bersentimen *compliant* or *not compliant*.

2. METODOLOGI PENELITIAN

2.1. Pengertian

Twitter adalah sebuah media sosial dan layanan *microblogging* yang memungkinkan penggunanya untuk mengirimkan pesan *realtime* (pesan ini populer dengan sebutan *tweet*). Berikut ini adalah beberapa istilah yang dikenal dalam Twitter [3].

A. Mention

Mention adalah menyebut atau memanggil pengguna Twitter lain dalam sebuah *tweet*. *Mention* dilakukan dengan menuliskan „@“ diikuti dengan nama pengguna lain.

B. Hashtag

Hashtag digunakan untuk menandai sebuah topik pembicaraan di Twitter. Penulisan *hashtag* dimulai dengan tanda „#“ diikuti dengan topik yang sedang dibahas. *Hashtag* biasa digunakan untuk meningkatkan visibilitas *tweet* pengguna.

C. Emoticon

Emoticon adalah ekspresi wajah yang direpresentasikan dengan kombinasi antara huruf, tanda baca dan angka. Pengguna biasa menggunakan *emoticon* untuk mengekspresikan *mood* yang sedang mereka rasakan.

D. Trending topics.

Jika *hashtag* adalah Cara untuk menandai sebuah topik pembicaraan di Twitter, maka *trending topics* adalah kumpulan dari topik pembicaraan yang sangat populer di Twitter.

Beberapa teknik yang digunakan untuk mendapatkan data atau informasi dalam pemecahan masalah. Metode yang dilakukan tersebut antara lain:

- a) **Studi Kepustakaan**
Metode pengumpulan data dengan metode kepustakaan dilakukan dengan pengumpulan jurnal, literatur, paper, makalah, buku, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan metode *Support Vectore Machine (SVM)*.
- b) **Pengumpulan Data Tweet**
Data yang diperoleh merupakan sumber yang diambil secara langsung dari Twitter dan menggunakan aplikasi RapidMiner dengan *keyword* pencarian @Ovo_ID .
- c) **Perancangan**
Dalam proses ini ada beberapa hal yang dilakukan seperti: Desain alur pengambilan data, desain alur sistem dan metode, desain pemrograman.
- d) **Pengujian dan Evaluasi**
Pada tahap ini dilakukan pengujian sistem untuk mengetahui dimana letak kesalahan yang mungkin terjadi dan dapat ditanggulangi.
- e) **Penyusunan Laporan**
Pada tahap ini penyusunan laporan dibuat sebagai dokumentasi yang berfungsi untuk dapat mempermudah dipelajari dan dikembangkan oleh orang lain.

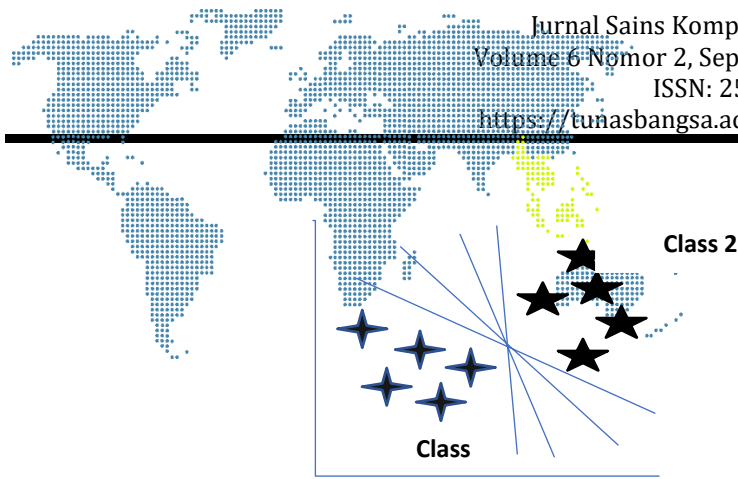
2.2. Metode *Support Vectore Machine*

Support Vector Machine Classification [4] adalah sistem pembelajaran mesin yang menggunakan ruang hipotetis berupa fungsi linier dalam ruang fitur berdimensi besar, dilatih dengan algoritma pembelajaran berdasarkan teori optimasi dengan mengimplementasikan deviasi pembelajaran yang dihasilkan dari teori pembelajaran statistik keluar.

Menurut [5], Support Vector Machine (SVM) adalah teknik prediksi yang relatif baru untuk klasifikasi dan regresi. Berada dalam kelas pembelajaran yang diawasi di mana perlu memiliki fase pelatihan menggunakan pelatihan SVM berurutan yang diikuti dengan fase pengujian.

Support Vector Machine (SVM) adalah metode klasifikasi menggunakan Machine Learning (supervised learning) yang memprediksi kelas dari model atau pola berdasarkan hasil proses pelatihan. Grading dilakukan dengan mencari hyperplane atau batas keputusan yang memisahkan satu kelas dengan yang lain, yang dalam hal ini berperan dalam memisahkan tweet sentimen positif (bertanda +1) dari tweet sentimen negatif (bertanda -1). SVM mencari nilai hyperplane menggunakan vektor bantu dan nilai margin. Pada penelitian ini data masukan yang memiliki representasi vektor diperoleh dari proses penimbangan. Dengan melakukan pelatihan dalam klasifikasi SVM, itu kemudian akan menghasilkan nilai atau pola yang akan digunakan dalam proses pengujian SVM untuk menandai sentimen di tweet.

Proses pembelajaran dalam masalah klasifikasi diterjemahkan sebagai upaya untuk menemukan garis (bidang-hiper) yang memisahkan kedua kelompok. Gambar tahapan pembelajaran di SVM adalah:



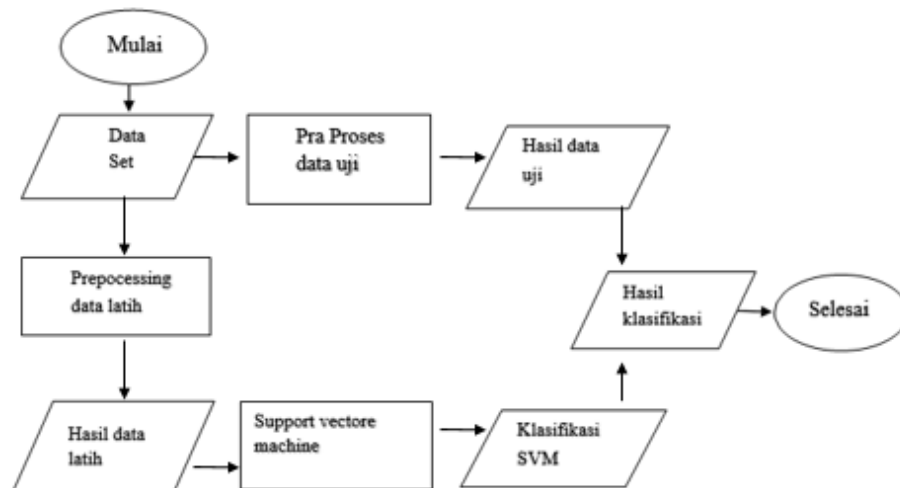
Gambar 1. *hyperplane* kelas Compliant dan Not Compliant

3. HASIL DAN PEMBAHASAN

3.1. Pembahasan

Pada bagian ini, penulis memakai Rapid Miner sebagai aplikasi dalam pembuatan analisis sentiment karna dalam pemakaian rapid miner simple dan mudah dipahami.

Tahapan dalam melakukan analisis sentimen dengan metode *klasifikasi SVM* dimulai dengan input data yang berupa data latih dan data uji yang kemudian diproses pada tahapan *pre-processing* hingga proses klasifikasi.



Gambar 2. Deskripsi Umum Alur Sistem

Gambar diatas menunjukkan tahapan-tahapan yang dijalankan pada penelitian ini. Berikut ini adalah penjabaran dari tiap-tiap proses yang dilakukan:

a) Input Data

Pengimputan berdasarkan data yang diambil dari twitter menggunakan rapidminer dan telah diberikan label pada setiap data ketika didalam excel.

b) Preproses program dan pembobotan kata

Data latih dan data uji yang telah diinputkan kemudian diproses oleh sistem untuk dilakukannya proses stopwords atau bisa disebut penghapusan kata tidak penting, contoh kata yang akan diproses oleh *stopword* seperti kata: yang,di,dan,kok,ke dll.

c) Penerapan Metode *Support Vector Machine*

Setelah penerapan preprocessing langkah selanjutnya masuk ke tahap metode *Support Vector Machine* yang akan melakukan pembobotan pada data yang telah diproses dan menentukan akurasi data latih yang telah di inputkan.

d) Proses pengimputan dan pengujian data uji

Masuk ketahap berikutnya, setelah proses pengujian data latih selanjutnya sistem yang dibuat akan melakukan pengujian menggunakan data uji yang di inputkan kedalam sistem untuk mendapatkan hasil sentimen dikelas Compliant atau Not Compliant yang diberi label (0) untuk Compliant dan (1) Not Compliant

Dataset dari hasil *crawling* ini akan dibagi menjadi dua bagian yaitu data *training* (data latihan) dan data *testing* (data uji) yang dipresentasikan pada Data latihan diklasifikasikan menggunakan *Support vectore machine* dengan label sentimen Compliant dan Not Compliant. Sedangkan data uji diklasifikasikan secara manual dengan label setimen compliant dan Not Compliant. Data uji nanti nya akan digunakan pada saat *evaluasi* untuk menentukan keakuratan data pada sistem. Pengambilan data *training* dan data *testing* dilakukan dengan waktu pengambilan yang berbeda.

Preprocessing adalah tahapan untuk mengolah kata mentah yang baru didapat. Langkah awal setelah data didapat adalah preprocessing. Data berupa kata-kata atau kalimat yang kita inputkan telah melalui tahap *filtering* dan *stopword* selanjutnya akan dilakukan pembobotan pada kata-kata tersebut setelah pembobotan diklasifikasi apabila nilai *term* pada kalimat bernilai >0 maka akan diletakan pada komentar positif dan apabila *term* kalimat bernilai < 0 maka akan diletakan pada komentar negatif.

Dalam klasifikasi *system* hanya melihat pada titik dan ruang pada dokumen tersebut untuk tujuan pemodelan ruang *vector* yang digunakan kemudian memberikan setiap kata dalam dokumen yang akan diproses. *Svm* dalam proses klasifikasi ini bertujuan untuk menemukan garis terbaik untuk membagi kedalam dua kelas kemudian diklasifikasikan dokumen uji berdasarkan pada sisi mana garis itu muncul. Apabila data yang dilatih terpisah secara linear maka bisa dilakukan pemilihan untuk dua *hyperplane* dari margin dengan cara menghilangkan atau tidak ada nilai antara *hyperplane* dan *margin* kemudian dimaksimalkan jaraknya dengan menggunakan persamaan $||^2||$. Cara ini untuk w meminimalkan $||w||$.

3.2. Pengujian

Pada proses pelatihan ini bertujuan pembentuk model prediksi data yang telah diambil. Untuk penelitian ini penulis menggunakan data latih berjumlah 1000 data yang diambil dari twitter dan dibagi menjadi dua model kelas yang pertama kelas "Complaint" dan yang kedua kelas "No Compliant". Berikut tampilan dari data latihan yang ditampilkan oleh sistem.

Pada penelitian ini menggunakan bilangan bulat karna pada perhitungan ini menentukan *range* dengan nilai 1-100 maka dikalikan

dengan 100, sehingga hasil yang didapat adalah 91. Sehingga hasil yang diklasifikasikan dengan tepat oleh sistem memiliki akurasi yang baik.

accuracy: 73.50% +/- 3.50% (micro average: 73.50%)

	true Complaint	true Not_Complaint	class precision
pred. Complaint	322	29	91.74%
pred. Not_Complaint	236	413	63.64%
class recall	57.71%	93.44%	

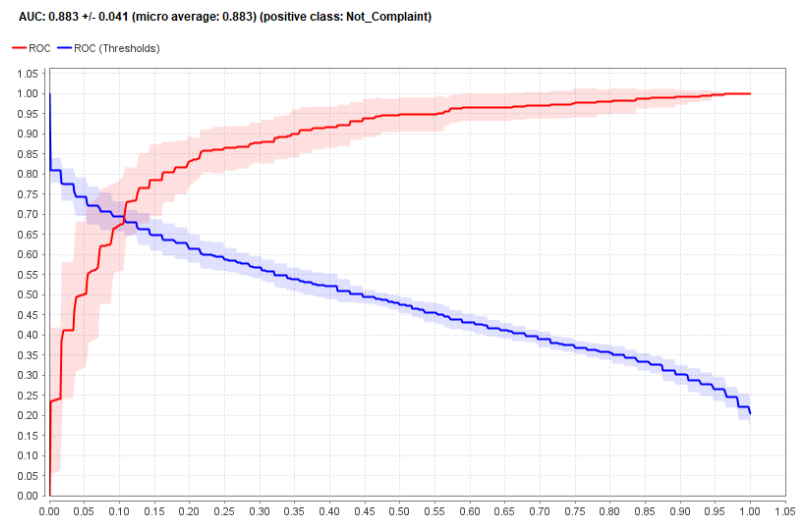
recall: 93.45% +/- 5.09% (micro average: 93.44%) (positive class: Not_Complaint)

	true Complaint	true Not_Complaint	class precision
pred. Complaint	322	29	91.74%
pred. Not_Complaint	236	413	63.64%
class recall	57.71%	93.44%	

precision: 63.69% +/- 2.97% (micro average: 63.64%) (positive class: Not_Complaint)

	true Complaint	true Not_Complaint	class precision
pred. Complaint	322	29	91.74%
pred. Not_Complaint	236	413	63.64%
class recall	57.71%	93.44%	

Gambar 3. Nilai akurasi, *recall* dan *precision*



Gambar 4. Grafik AUC

4. SIMPULAN

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa penelitian tentang analisis sentimen pengguna Twitter pada topik #BPJSrasarentenir dengan 1000 sample data menggunakan metode *Support Vectore Machine* telah berhasil dilakukan, denga hasil akurasi 96%. Penelitian ini menghasilkan model yang dapat melakukan klasifikasi dan prediksi tentang tanggapan masyarakat terhadap Ovo_ID apakah itu

tanggapan compliant maupun not compliant. Dan pada penelitian ini analisis sentiment masyarakat tentang #Ovo_ID lebih beranggapan not compliant.

DAFTAR PUSTAKA

- [1] Utomo, M. S, 2013, "Implementasi Stemmer Tala pada Aplikasi Berbasis Web", Jurnal Teknologi informasi DINAMIK, Volume 18, pp. , 41-45.
- [2] Dian Indriani, 2019., "Analisis Sentiment pada tweet dengan tagar #kpujangancurang menggunakan metode Naïve Bayes", Riau, Universitas Islam Riau.
- [3] Idris, N.R.N., Toh, C.L., and Elbuluk, E., "A New Torque and Flux Controller for DTC of Induction Machine", IEEE Transactions on Industry Application, Vol. 42, No.6, pp. 1358-1366, Nov/Dec 2006.
- [4] M. Hosoz and H. M. Ertunc, "Artificial Neural Network Analysis of an Automobile Air Conditioning System" Energy Conv. Man., Vol. 47, pp. 1574-1587, July 2006.
- [5] B. Liu, 2012. "Sentiment Analysis and opinion mining". Morgan & Claypool, Publisher.
- [6] Mark Ryan M. Talabis. D. Kaye, 2015. "Supervised learning", Nugroho, A. S. 2008. "Support Vector Machine, Paradigma baru dalam softcomputing", Konferensi Nasional Sistem dan Informatika, 92-99.
- [7] Mhd. Furqan, dkk. 2019. "Penerapan Metode Otsu dalam melakukan Segmentasi Citra pada Citra Naskah Arab", Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, November, ISSN 2476-9843.20(1):59-72.
- [8] Retno Tri Wulandari, 2017, "Data Mining". Yogyakarta: Penerbit Gava Media. Hal 33-34.
- [9] Novianto Donna Prayoga, dkk. 2018, "Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Agustus, ISSN: 2548-964X, 2666-2671.
- [10] Bukhari, Varian Habibie. (2015). "Sentiment Analysis Menggunakan K-Nearest Neighbor dengan perbandingan Fungsi jarak (Studi Kasus : Twitter Indosat dan Telkomsel)". Bandung : Universitas Widyatama.