

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN UNIT GAWAT DARURAT RUMAH SAKIT UMUM DI INDONESIA MENGGUNAKAN SELEKSI FITUR INFORMATION GAIN DAN SUPPORT VECTOR MACHINE

Reja Saptari¹, Rianto², Acep Irham Gufroni³.

Program Studi Informatika Fakultas Teknik Universitas Siliwangi Tasikmalaya
Reja.saptari14@student.unsil.ac.id¹, rianto@unsil.ac.id², acep@unsil.ac.id³

ABSTRAK

Sentiment analisis adalah proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial fokus utamanya adalah menganalisa dokumen teks untuk memperoleh opini dari pengguna yang terdapat pada beragam platform media sosial. Pada penelitian ini dilakukan analisis sentiment postingan opini masyarakat terhadap pelayanan unit gawat darurat rumah sakit umum di Indonesia pada twitter yang nantinya menghasilkan informasi opini sentimen masyarakat berupa sentimen positif, negatif dan netral. Dalam penelitian ini dilakukan seleksi fitur menggunakan Information Gain dan Support Vector Machine sebagai algoritma klasifikasi text untuk mengklasifikasi sentiment opini masyarakat pada tweet twitter. Pada penelitian ini menunjukkan nilai Prediksi yang didapat dari hasil Implementasi metode Support Vector Machine dengan menggunakan Seleksi Fitur Information Gain adalah AUC 0.993, CA 0.980, F1 0.980, Precision 0.981 dan Recall 0.980.

Kata kunci: Seleksi fitur, Information gain, SVM, Analisis sentiment, Unit gawat darurat, *twitter*

ABSTRACT

Sentiment analysis is the process of using text analytics to obtain various data sources from the internet and various social media platforms, the main focus of which is analyzing text documents to obtain opinions from users on various social media platforms. In this study, sentiment analysis was carried out by posting public opinion on the services of the emergency unit of general hospitals in Indonesia on Twitter which later produced information on public sentiment opinions in the form of positive, negative and neutral sentiments. In this study, feature selection was conducted using Information Gain and Support Vector Machine as a text classification algorithm to classify public opinion sentiment on Twitter tweets. In this study, the prediction values obtained from the implementation of the Support Vector Machine method using the Information Gain Feature Selection are AUC 0.993, CA 0.980, F1 0.980, Precision 0.981 and Recall 0.980.

Keywords: Feature selection, Information gain, SVM, Sentiment analysis, Emergency department, twitter

PENDAHULUAN

Sosial media merupakan suatu media yang dapat digunakan untuk berekspresi oleh penggunanya. Sosial media Twitter cukup populer dan sering digunakan di Indonesia, pengguna twitter dapat berekspresi dan berinspirasi tanpa adanya batasan. Sosial

Media khususnya Twitter merupakan salah satu platform sosial media yang sering digunakan masyarakat Indonesia. Sosial media twitter merupakan media sosial yang fokus utamanya pada teks atau lebih dikenal dengan sebutan 'tweet', tweet ini lah yang menjadi point unggulan Sosial Media Twitter

dimana setiap penggunaanya bebas untuk mengeluarkan isi pemikiran dan hatinya pada tweet- tweet nya. Sosial media Twitter merupakan suatu media yang dapat digunakan untuk ulasan sebuah produk atau layanan oleh penggunaanya, khususnya pelayanan unit gawat darurat pada rumah sakit di Indonesia.

Berdasarkan UU RI Nomor 44 Tahun 2009, pelayanan kesehatan paripurna yang dimaksud adalah pelayanan kesehatan yang meliputi *promotif, preventif, kuratif* dan *rehabilitatif*. Pelayanan tugas kesehatan perorangan secara paripurna tersebut, pada dasarnya rumah sakit mempunyai fungsi menyelenggarakan pelayanan pengobatan dan pemulihan kesehatan sesuai dengan standar pelayanan rumah sakit.

Pada penelitian (Fathan Hidayatullah dan Sn, 2014) mengenai analisis sentiment dan klasifikasi kategori terhadap tokoh publik pada twitter yang mengkomparasi Algoritma Klasifikasi antara *Support Vector Machine* dengan *Naïve bayes* menghasilkan hasil bahwa Algoritma *Support Vector Machine* menghasilkan akurasi performansi yang lebih baik dibandingkan dengan Algoritma *Naïve bayes* dalam hal Klasifikasi Teks. Pada penelitian yang dilakukan oleh (Siti Nur Asiyah, 2016) mengenai komparasi Algoritma Klasifikasi teks antara *Support Vector Machine* dengan *K-Nearest Neighbor* untuk klasifikasi berita online menunjukkan bahwa performansi *Support Vector Machine* jauh lebih baik daripada *K-Nearest Neighbor* dalam hal klasifikasi teks. Menurut (Maulana dan Soebroto, 2019) Seleksi fitur adalah proses untuk membuang fitur-fitur yang tidak atau

kurang relevan pada proses klasifikasi text. Penelitian yang dilakukan oleh (Sari, 2016) Pada penelitian mengenai implementasi seleksi fitur *Information Gain* pada algoritma klasifikasi *machine learning* untuk prediksi performa akademik siswa, penelitian ini menunjukkan bahwa dengan implementasi seleksi fitur *Information gain* dapat mempengaruhi hasil dari algoritma klasifikasi *Machine learning (J48, Random Forest, MLP, SVM (SMO), dan Naïve Bayes)*.

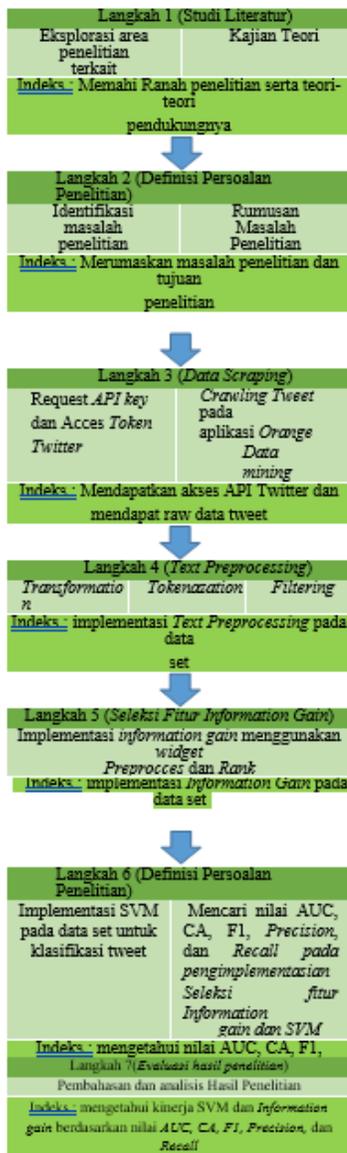
Pada penelitian yang dilakukan oleh (Rosdiana *et al.*, 2019) mengenai Dengan kondisi tersebut dapat dilakukan analisis sentimen publik terhadap pelayanan pemerintah Kota Makassar menggunakan data twitter. Analisis data yang dihasilkan dapat menunjukkan sentimen positif, sentimen negatif maupun sentimen netral masyarakat menggunakan metode klasifikasi *Naïve Bayes Classifier (NBC)* yang nantinya dapat digunakan oleh pemerintah Kota Makassar untuk meningkatkan kualitas pelayanan terhadap masyarakat. Pada penelitian ini pun memiliki topik yang hampir sama yaitu mengenai pelayanan.

Berdasarkan seluruh uraian tersebut, penelitian ini akan membahas tentang “analisis sentimen pengguna twitter terhadap pelayanan unit gawat darurat rumah sakit di indonesia menggunakan seleksi fitur *information gain* dan *support vector machine*”.

METODE PENELITIAN

Pada penelitian ini diawali dengan Studi pustaka mempelajari tentang teori-teori yang menjadi referensi dan pendukung dalam pembuatan dan penulisan hasil penelitian mengenai analisis sentimen

pelayanan unit gawat darurat terhadap rumah sakit di Indonesia berdasarkan tweet pada pengguna twitter dengan menggunakan klasifikasi information gain dan algoritma Support Vector Machine (SVM). Pada tahap pengumpulan data tweet pada pengguna twitter dilakukan dengan menggunakan twitter API yang telah disediakan oleh twitter untuk tujuan penelitian dan pembelajaran. Pada proses pengambilan data dan analisis sentimen ini akan menggunakan aplikasi orange data mining untuk memudahkan proses pengambilan data.



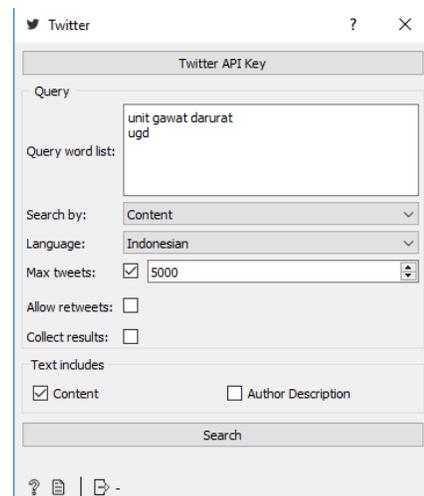
HASIL DAN PEMBAHASAN

Data Mining

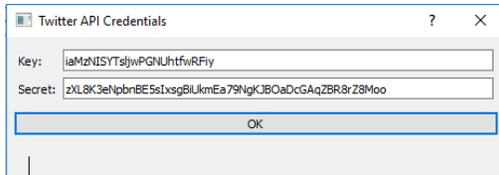
Setelah mendapatkan akses dari API twitter, langkah selanjutnya adalah membuat dataset. Untuk membuat dataset ini digunakan aplikasi *Orange Data Mining*, aplikasi ini merupakan aplikasi atau *platform data analytics* yang menggunakan konsep *Visual Programing (Free Coding)*. Dalam aplikasi ini terdapat beberapa *widget* yang digunakan untuk membuat data set mengenai tweet yang telah dicrawling sebelumnya.

Widget Twitter

Widget twitter ini digunakan untuk mengakses data tweet dari twitter. Pada menu *Twitter API key* ini diharuskan untuk mengisi *API key* dan *Access token* sebagai autentifikasi untuk mengakses data dari twitter. Kemudian pada *Query word list* di isikan kata kata yang berkaitan dengan unit gawat darurat. Dikarenakan pada penelitian ini yang dicari adalah tweet unit gawat darurat maka pencarian dilakukan berdasarkan konten.



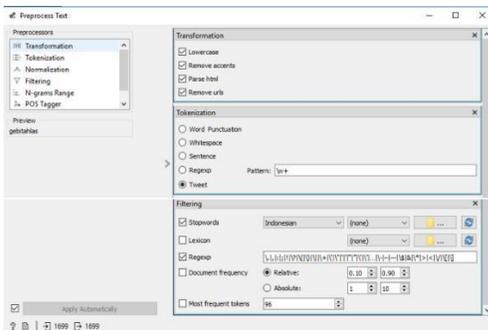
Gambar 1. *Widget Twitter*



Gambar 2. Widget Twitter (Twitter API Key)

Widget Preprocess Text

Widget preprocess text, widget ini dapat membagi teks menjadi sebuah unit-unit yang lebih kecil (token), memfilternya, menjalankan normalisasi (stemming, lemmatization), membuat n-grams dan memberi tag token dengan label part-of-speech. Pada widget preprocess text terdapat 6 fitur preprocessors yang dapat digunakan yaitu, Transformation, Tokenization, Normalization, Filtering, N-grams Range, dan POS Tagger.



Gambar 3. Widget Preprocess Text

Perujukan pustaka mengikuti aturan perujukan. Jika penulis hanya seorang ditulis seperti ini (Amidi, 2017), jika penulis ada dua orang maka ditulis seperti ini (Amidi & Andi, 2016). Jika penulis lebih dari 3 maka cukup ditulis penulis pertama ditambah *et al.* seperti Apriliani *et al.* (2017). Setiap pustaka yang dirujuk dalam artikel ini harus tercantum dalam daftar pustaka.

Pada penelitian kali ini fitur yang digunakan adalah ;

1. Transformation

- a) *Lower case* : fitur ini digunakan untuk merubah data masukan

dan mentranformasi data masukan tersebut menjadi huruf kecil.

- b) *Remove accents* : fitur ini digunakan untuk menghapus semua aksen atau diacritics pada teks (*naïve > naive*).
 - c) *Parse HTML* : fitur ini digunakan untuk mendeteksi tag html dan merubahnya menjadi kalimat biasa.
 - d) *Remove URLs* : fitur ini digunakan untuk menghapus url.
2. *Tokenazation* adalah metode yang digunakan untuk memecah teks menjadi komponen yang lebih kecil.

- a) *Whitespace* : fitur ini digunakan untuk memisahkan suatu kalimat menjadi kata-kata berdasarkan spasi.

- b) *Regexp* : fitur ini digunakan untuk memisahkan teks berdasarkan ekspresi regular yang ada pada aplikasi *Orange Data mining*.

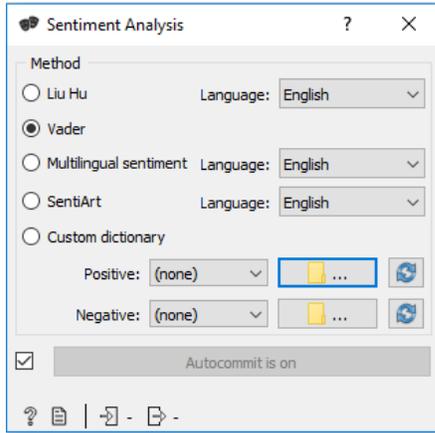
3. Filtering

- a) *Stopwords* : fitur ini digunakan untuk menghapus stopword dari teks. Pada aplikasi *Orange data mining* ini telah disediakan pula *library stopwords* untuk Bahasa Indonesia ataupun kita dapat menginput sendiri *stopwords* yang akan kita gunakan.
- b) *Regexp symbols*: fitur ini digunakan untuk menghapus tanda baca.

Widget Sentiment Analysis

Widget sentiment analysis memprediksi sentiment untuk setiap data. widget sentiment analysis menyediakan 5 method sentiment yaitu Liu Hu, Vader, Multilingual sentiment, SentiArt, dan Custom dictionary dari method ke 5 tersebut

adalah algoritma lexicon-based. Widget sentiment analysis akan menambahkan 4 fitur baru dari Vader method: positive score, negative score, neutral score dan compound (combined score).



Gambar 4. *Widget Sentiment Analysis*

Penerapan *Information Gain* pada *Orange Data Mining*

Penerapan seleksi fitur *information gain* pada aplikasi orange data mining pada suatu data set, yaitu dengan menggunakan *widget Rank* yang ditunjukkan pada gambar 4.9. Pada *widget rank*, *information gain* dapat diterapkan dengan memilih pada menu *scoring methods*. Kelebihan pada *widget rank* ini hasil dari perhitungan seleksi fitur *information gain* dapat di tampilkan dalam bentuk tabel dan diurutkan berdasarkan hasil perhitungan seleksi fitur *information gain* tersebut. Dari hasil perhitungan seleksi fitur *information gain* ada 4 atributtes nilai tertinggi yang di pilih yaitu *compound*, netral, positif, dan negatif. *Widget* yang digunakan selanjutnya setelah menerapkan seleksi fitur *information gain* adalah *widget data sampler* Penerapan *Support Vector Machine(SVM) Widget SVM* terdapat dua tipe SVM yang dapat digunakan yaitu SVM dan v-SVM yang didasarkan pada minimalisasi fungsi kesalahan yang berbeda. Pada saat memilah

tipe SVM dapat ditentukan nilai *Cost* (C) dan *Regression loss epsilon* (ϵ), dan jika memilih v-SVM maka dapat di tentukan nilai *Cost* (c) dan *Complexity Bound* (v) . *Cost* sendiri merupakan penalty untuk *loss* dan berlaku untuk klasifikasi dan tugas regresi. Kemudian *Regression loss epsilon* (ϵ) merupakan parameter untuk model *epsilon-SVR*, dan berlaku untuk tugas regresi. Kemudian mendefinisikan jarak dari nilai sebenarnya di mana tidak ada penalti yang dikaitkan dengan nilai prediksi. *Complexity bound* (v) merupakan parameter untuk model v-SVR, berlaku untuk tugas klasifikasi dan regresi. Batas atas pecahan kesalahan pelatihan dan batas bawah pecahan vektor pendukung. SVM Type yang digunakan pada proses ini adalah SVM. Untuk kernel yang akan digunakan adalah kernel Linear. Kernel adalah fungsi yang mengubah ruang atribut menjadi ruang fitur baru agar sesuai dengan hyperplane margin maksimum, sehingga memungkinkan algoritma untuk membuat model dengan kernel Linear, Polynomial, RBF, dan Sigmoid. Kemudian dapat ditetapkan nilai penyimpangan dari nilai yang diharapkan dalam Toleransi Numerik. Terdapat pula pilihan untuk menggunakan Batas Iterasi, Batas iterasi ini digunakan untuk mengatur jumlah maksimum iterasi yang diizinkan. Hasil output dari widget SVM selanjutnya di input pada widget *predictions* sehingga widget *prediction* ini menerima dua input yaitu Input langsung dari widget data sampler sebagai data remaining yang masuk sebesar 30% dan input dari widget SVM sebagai data sample sebesar 70%.

Nilai prediksi yang di dapat dari penelitian dengan hasil Implementasi metode *Support Vector Machine* dengan menggunakan seleksi Fitur *Information gain* adalah AUC 0.993, CA 0,980, F1 0.980, *Precision* 0.981 dan *Recall* 0.980. Sebagai

nilai pembandingan. sedangkan jika tanpa menggunakan seleksi fitur *Information Gain* akan diperoleh hasil *AUC 0.553, CA 0.811, F1 0.727, Precision 0.658, dan Recall 0.811.*

PENUTUP

Berdasarkan hasil pengujian dan analisa maka dapat disimpulkan :

1. Pada penelitian ini berhasil menerapkan seleksi fitur *Information Gain* pada data set tentang unit gawat darurat rumah sakit di indonesia yang didapat dari twitter. Data yang didapatkan kemudian dilakukan juga klasifikasi sentiment data terhadap unit gawat darurat rumah sakit di indonesia pada dataset yang sudah dilakukan seleksi fitur *Information Gain* dengan menggunakan metode klasifikasi *Support Vector Machine*.
2. Nilai Prediksi yang didapat dari hasil Implementasi metode *Support Vector Machine* dengan menggunakan Seleksi Fitur *Information gain* adalah *AUC 0.993, CA 0.980, F1 0.980, Precision 0.981 dan Recall 0.980.* Sebagai nilai pembandingan sebelum menggunakan Seleksi Fitur *Information gain* adalah *AUC 0.553, CA 0.811, F1 0.727, Precision 0.658, dan Recall 0.811.* Hal ini menunjukkan bahwa seleksi fitur menggunakan *Information gain* berpengaruh pada nilai prediksi yang dihasilkan. Hasil klasifikasi sentimen pengguna twitter terhadap unit gawat darurat rumah sakit di indonesia menggunakan Algoritma *Support Vector Machine* pada confusion matrix mendapatkan hasil prediksi nilai positif sebesar 38 tweet, netral sebesar 418 tweet sebesar dan negatif sebesar 53 tweet. Hasil persentase positif dan negatif dapat berubah sewaktu-waktu sesuai dengan data set dan opini yang diberikan masyarakat terhadap unit gawat darurat rumah sakit di indonesia.

DAFTAR PUSTAKA

- Al Anfal (2020) “Pengaruh Kualitas Pelayanan Dan Citra Rumah Sakit Terhadap Tingkat Kepuasan Pasien Rawat Inap Rumah Sakit Umum Sundari Medan Tahun 2018,” *Journal Of Chemical Information and Modeling*, 53(9), hal. 1689–1699.
- Asiyah, S. N. dan Fithriasari, K. (2016) “Klasifikasi Berita Online Menggunakan Metode *Support Vector Machine* Dan *K-Nearest Neighbor*,” *Jurnal Sains dan Seni ITS*, 5(2), hal. 317–322. doi: 10.12962/j23373520.v5i2.16643.
- Fathan Hidayatullah, A. dan Sn, A. (2014) “ISSN: 1979- 2328 UPN "Veteran,” Seminar Nasional Informatika, 2014(semnasIF), hal. 115–122.
- Manomaisupat, P. dan Ahmad, K. (2005) “Feature selection for text categorisation using self- organising map,” *Proceedings of 2005 International Conference on Neural Networks and Brain Proceedings, ICNNB’05*, 3(June), hal. 1875–1880. doi: 10.1109/icnnb.2005.1614991.
- Maulana, M. I. dan Soebroto, A. A. (2019) “Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode *Improved k- Nearest Neighbor* dan Seleksi Fitur *Chi- square*,” 3(7), hal. 6662–6669.
- Maulana, M. R. dan Karomi, M. A. Al (2015) “*Information Gain* Untuk Mengetahui Pengaruh Atribut,” *Jurnal Litbang Kota Pekalongan*, 9, hal. 113–123.

- Maulida, I. et al. (2016) “Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain,” Oktober 2016 Ijccs, 17(2), hal. 1–5.
- Rosdiana et al. (2019) “Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar,” Proceeding SNTEI, hal. 87–93.
- Pratama, A., Wihandika, R. C. dan Ratnawati, D. E. (2018) “Implementasi algoritme support vector machine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2(April), hal. 1704–1708.
- Ratniasih, N. L., Sudarma, M. dan Gunantara, N. (2017) “Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat,” Majalah Ilmiah Teknologi Elektro, 16(3), hal. 13. doi: 10.24843/mite.2017.v16i03p03.
- Sari, B. N. (2016) “Implementasi Teknik Seleksi Fitur Information Gain pada Algoritma Klasifikasi Machine Learning untuk Prediksi Performa Akademik Siswa,” Seminar Nasional Teknologi Informasi dan Multimedia 2016, (March), hal. 6– 7.
- Simanjuntak, steven roy (2017) “Text Mining Untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes (NB),” Jurnal Telematika, 12(01).
- Syafaruddin dan Syukur, A. T. (2014) “Analisis Kualitas Pelayanan Kesehatan Pada Unit Gawat Darurat Rumah Sakit Umum Daerah Haji Makassar,” 20, hal. 52–61.
- Utami, L. D. (2015) “Integrasi Metode Information Gain untuk Seleksi Fitur dan AdaBoost untuk Mengurangi Bias pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naive Bayes,” Journal of Intelligent Systems, 1(2), hal. 120–126.