

Analisa Sentimen Review Restoran di Situs *Microblog Twitter* Menggunakan Algoritma *Support Vector Machine*

Mochammad Aldin¹, Ermatita²

^{1,2}Fakultas Ilmu Komputer

¹Universitas Pembangunan Nasional “Veteran” Jakarta

²Universitas Sriwijaya

¹aldinmoehammad1892@gmail.com

Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

Abstrak

Perkembangan situs microblog saat ini sangatlah pesat, dari yang sebelumnya hanya menjadi sarana berbagi kata antar pengguna, sekarang sudah menjadi sebuah keharusan di bidang bisnis sebagai sarana periklanan. Salah satunya adalah restoran, saat ini hampir seluruh restoran menggunakan media sosial sebagai sarana pemasaran dan juga untuk berkomunikasi dengan pelanggan. Sehingga pelanggan juga bebas menulis sebuah komentar yang ditujukan sebagai kritik maupun saran terhadap restoran tersebut atau biasa disebut dengan review atau ulasan. Pada aplikasi ini tweet diklasifikasikan menjadi dua nilai kelas yang mana kelas itu adalah kelas bernilai positif atau negatif. Algoritma dalam penelitian ini adalah Support Vector Machine (SVM), algoritma ini digunakan untuk proses klasifikasi sentiment pada tweet. Data didapat dengan menggunakan API (Application Programming Interface) yang disediakan oleh pihak Twitter. Sehingga didapatkan 414 tweet sebagai data latih. Algoritma Support Vector Machine (SVM) mendapatkan hasil pengujian dengan akurasi 82,92% dengan nilai precision sebesar 82,92% dan nilai recall 83%. Sehingga menunjukkan bahwa algoritma Support Vector Machine bisa digunakan dalam melakukan klasifikasi terhadap sentimen positif dan negatif terhadap review restoran.

Kata Kunci: *Tweet, Review, Klasifikasi, Support Vector Machine, Praproses*

1 PENDAHULUAN

Media sosial sudah menjadi makanan sehari – hari pada masa ini, Beragamnya konten dan informasi yang disediakan menjadi daya tarik tersendiri bagi para penggunanya. Bagi para pelaku bisnis khususnya di bidang kuliner, media sosial sudah menjadi kebutuhan wajib sebagai sarana pemasaran mereka. Hal ini juga menjadi keuntungan bagi para pengguna lainnya seperti pelanggan yang telah berkunjung ke restoran tersebut maupun calon pelanggan yang mencari rekomendasi restoran. Selain mendapatkan informasi tentang produk dan promo, mereka juga dapat menulis opini yang ditujukan untuk memberikan kritik maupun saran bagi para pemilik restoran yang biasa disebut dengan istilah review atau ulasan.

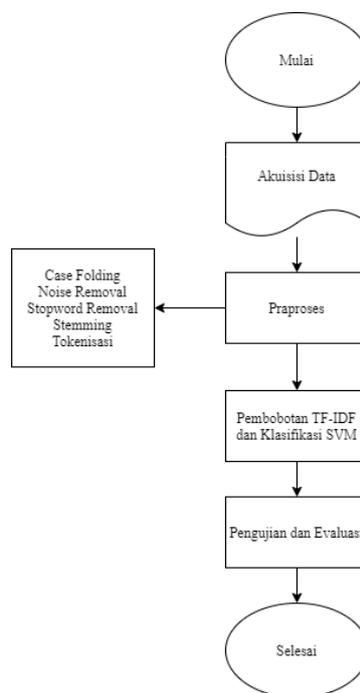
Dari ulasan inilah nilai sebuah restoran dapat terlihat, baik bernilai positif maupun negatif. Namun dengan banyaknya ulasan yang ada juga penggunaan kata yang bersifat opini, para pengguna lain menjadi kesulitan untuk mendapatkan makna dan nilai sebenarnya dari ulasan tersebut, sehingga dapat mengurangi akurasi nilai.

Dalam melakukan analisis sentimen diperlukan metode yang menunjang klasifikasi dan Support Vector Machine (SVM) dipilih menjadi metode yang digunakan. Support Vector Machine dipakai sebagai metode klasifikasi ulasan yang memisahkan kelas sentimen dan mengkategorikannya ke dalam dua nilai kelas yaitu kelas sentimen positif dan kelas sentimen negatif, SVM digunakan juga untuk menambah akurasi dalam proses klasifikasi tersebut. Sehingga pengguna mendapatkan informasi ulasan yang telah bernilai positif atau negatif juga memiliki akurasi klasifikasi yang tinggi. Support Vector Machine (SVM) dinilai sangat cocok dan akurat untuk melakukan klasifikasi dua kelas dalam penelitian

ini. Hal ini dikarenakan pada SVM klasifikasi dilakukan dengan cara mencari hyperlane atau garis pemisah diantara dua kelas tersebut maka hasil yang dicapai akan lebih akurat jika dibandingkan dengan metode lainnya. Jika pada algoritma Naïve Bayes memiliki kemudahan terhadap implementasinya, Key Nearest Neighbor dapat meminimalkan resiko empiris, namun SVM dipilih karena algoritma ini dapat berfungsi secara optimal pada data yang memiliki dimensi tinggi jadi dinilai cocok untuk klasifikasi terhadap analisis sentimen. Pada SVM juga terdapat beberapa jenis kernel, kernel inilah yang menjadikan nilai akurasi lebih meningkat dan dapat disesuaikan terhadap data yang ada

2 METODOLOGI PENELITIAN

2.1 Akuisisi Data



Gambar 1: Kerangka pikir di bawah berisi proses yang telah dilakukan dalam pengerjaan penelitian

Tahap ini merupakan tahap untuk mendapatkan data yang dibutuhkan, data yang dimaksud dalam penelitian ini adalah tweet pengguna twitter yang mention akun official @Mcdonalds_ID yang selanjutnya digunakan untuk data latih dan data uji. Data tweet didapatkan dengan cara menggunakan API(Application Programming Interface) yang sudah disediakan oleh twitter sebanyak 414 data tweet.

2.2 Preprocessing

Untuk menghilangkan noise dari data yang akan diolah dilakukan tahap ini juga untuk mengurangi attribute pada data yang kurang berpengaruh. Sehingga memudahkan dan memaksimalkan proses klasifikasi.

2.3 Pembobotan Kata TF-IDF

Pembobotan kata adalah proses pemberian bobot nilai pada tiap kata yang ada terhadap data latih maupun data uji. Sehingga mempermudah untuk tahap klasifikasi. Pada penelitian ini pembobotan menggunakan Term Frequency – Inverse Document Frequency.

2.4 Klasifikasi Dengan Support Vector Machine

Klasifikasi merupakan tahap akhir dari penelitian ini yaitu pengklasifikasian data tweet yang telah dibersihkan datanya dan telah dilakukan tahap praproses kemudian klasifikasi menggunakan Support Vector Machine dilakukan. Algoritma tersebut digunakan guna mencari ulasan yang mengandung nilai positif dan nilai negatif. Dalam pengerjaan menggunakan Support Vector Machine data dibagi menjadi dua, yaitu data training dan data testing. Pembagian data ini menggunakan Cross Validation Satisfied K-fold yang membagi data menjadi data uji dan data latih. Dalam melakukan pencarian dua kelas tersebut Support Vector Machine bekerja dengan mencari garis pemisah dari dua kelas tersebut atau biasa disebut dengan istilah hyperlane. Hyperlane merupakan pemisah antara dua kelas yang dapat ditemukan dengan cara mengukur margin yang dapat ditemukan dengan cara mengukur margin yang dapat ditemukan dengan cara mengukur margin pada hyperlane dan mencari titik dengan nilai terbesarnya. Margin merupakan jarak antara hyperplane tersebut dengan data yang terdekat dari masing-masing kelas. Klasifikasi yang dilakukan menggunakan kernel Radial Basis Function (RBF).

2.5 Pengujian dan Evaluasi

Evaluasi dilakukan untuk mengukur kinerja klasifikasi yang dijalankan oleh program yang saya buat, bila nilai akurasi kurang dari 80% maka akan ditambahkan proses pada tahap preprocessing data sehingga mendapatkan 80% atau lebih.

3 HASIL DAN PEMBAHASAN

Penelitian ini memiliki dan menggunakan data yang merupakan semua tweet yang memuat username @McDonalds_ID. Hasil proses crawling pada penelitian ini sebanyak 414 tweet yang selanjutnya diberi label tweet positif dan negatif.

Data kemudian melalui tahap praproses dimana semua noise dari data tersebut dihilangkan

Tabel 1: Tabel sampel data latih

Sampel Data Latih	
Data Latih	Label
Terima kasih @McDonalds_ID buat sambal bawangnya. gila enak banget..!!!! Terlihat enak, yuk	Positif
@McDonalds_ID Bikin gini jugaa https://t.co/eftB1a7sU3 Hey mekdi @McDonalds_ID	Positif
saya sebagai pelangganmu kecewa sekali!!	Negatif
aku sebagai pecinta mcd sangat kecewa ❤️ @McDonalds_ID	Negatif

Tabel III. Hasil Akhir Praproses

Tweet Asli	Hasil Akhir
dapet menu murah porsi mini pas gak gak pesan langsung buka puasa sahur terimakasih	['dapet', 'menu', 'murah', 'porsi', 'mini', 'pas', 'pesan', 'langsung', 'buka', 'puasa', 'sahur', 'terimakasih']

Data pada tahapan preprocessing memiliki hasil akhir yang merupakan token atau kata - kata yang telah dipisah dari sebuah kalimat. Namun pada hasil akhir preprocessing setiap kata belum memiliki nilai atau bobot maka dibutuhkanlah proses pembobotan kata. Pembobotan kata atau term weighting pada penelitian ini dilakukan dengan metode Term Frequency – Inverse Document Frequency (TF-IDF). Dengan metode ini kata – kata yang terdapat pada tweet memperoleh nilai fitur yang dapat selanjutnya digunakan dalam tahapan klasifikasi sentimen juga dapat membantu menaikkan akurasi dari klasifikasi sentimen itu sendiri.

Setelah kata diberikan bobot atau nilai menggunakan Term Frequency – Document Inverse Frequency selanjutnya setiap nilai pada kata akan di akumulasi dan mulai dibagikan berdasarkan kelasnya. Dalam penelitian ini sentiment di analisis menggunakan algoritma Support Vector Machine (SVM) yang menggunakan library SCIKIT – Learn dan dihitung otomatis dengan kernel yang terdapat pada algoritma SVM yaitu kernel Radial Basis Function (RBF). Data latih dan data uji dibagi secara otomatis menggunakan StratifiedKFold dengan n-split 5 dan menggunakan 9 random state, hasilnya sistem membagi data latih menjadi 80% dari total hasil crawling sedangkan 20% lainnya merupakan data uji. Dalam penelitian ini peneliti telah mencoba beberapa kernel yaitu kernel Linear, Polynomial, Sigmoid dan Radial Basis Function.

Dapat disimpulkan bahwa nilai akurasi pada program ini sebesar 82,92 % dengan nilai recall sebesar 79,36% dengan nilai true positive sebanyak true positive 50, true negative 18, false positive 1 dan false negative 13. Besaran akurasi algoritma Support Vector Machine pada penelitian ini sebesar 82,92% sehingga dapat dikatakan bahwa algoritma ini dapat melakukan klasifikasi terhadap sentimen review restoran secara efektif.

4. KESIMPULAN

Penggunaan algoritma Support Vector Machine (SVM) dalam klasifikasi review sentimen tweet terhadap review restoran dapat dilakukan dengan baik. Algoritma Support Vector Machine dalam klasifikasi sentimen tweet terhadap review restoran berjalan dengan baik dan memiliki nilai akurasi sebesar 82,92% dengan dan nilai recall 83%.

Referensi

- Hadi, Mulya., 2010. Twitter untuk Orang Awam. Palembang. Maxikom.
- Liu, B., 2010, Sentiment Analysis Multi Facet Problem. In 2010 Conference for IEEE IntelligentSystem.
- Utomo, M. S., 2013. Implementasi Stemmer Tala pada Aplikasi Berbasis Web. Jurnal Teknologi Informasi DINAMIK.
- Hidayat, A. (2016). Algoritma TF-IDF, dari Arfian Hidayat: <https://arfianhidayat.com/algoritma-tf-idf>.
- S. Fachrurrazi., 2012. Penggunaan Metode Support Vector Machine (SVM) Untuk Mengklasifikasi dan Memprediksi Angkutan Udara dan Jenis Penerbangan Domestik dan Internasional di Banda Aceh. TECHSI-Jurnal Teknik Informatika.
- Richardson, M dan S.Wallace. 2013 Getting Started With Raspberry Pi. O'Reilly Media, Inc.,USA.