

PLAGIARISM SCAN REPORT

Words 896 Date January 08,2020

Characters 6820 Exclude Url

5%

Plagiarism

95%

Unique

2

Plagiarized
Sentences

39

Unique Sentences

Content Checked For Plagiarism

a. Transform Cases Dalam proses ini, kata-kata yang tidak relevan akan diubah, seperti kata yang mengandung huruf besar yang diubah menjadi huruf kecil sehingga dapat saling berhubungan dengan sentiment. Tabel 6. Proses Transform Cases Sebelum Transform Case Setelah Transform Case Utk mengganti presiden kita harus TENGSELAMKAN BANTENG di setiap Daerah, GANTIPRESIDEN utk mengganti presiden kita harus tenggelamkan banteng di setiap daerah, gantipresiden Sumber: Hasil Penelitian (2019) b. Generate n-Grams (bigram) Proses n-Grams digunakan untuk menyelesaikan permasalahan klasifikasi dokumen ke dalam sentiment positif atau negatif. Kesalahan klasifikasi sentiment biasanya disebabkan oleh fitur term tunggal. Misalnya kata "buruk" merupakan kata yang termasuk sentiment negatif, akan tetapi kata "tidak buruk" masuk ke dalam sentiment positif apabila berdampingan dengan kata negasi begitupun sebaliknya. 1. Klasifikasi **Text mining dapat menganalisis dokumen, mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta menentukan kesamaan di antara dokumen untuk mengetahui bagaimana mereka berhubungan dengan variabel lainnya. Text mining merupakan teknik yang digunakan untuk menangani masalah classification, clustering, information extraction dan information retrival (Berry & Kogan, 2013).** Validation adalah proses untuk mengevaluasi keakuratan prediksi dari model. Validasi digunakan untuk memperoleh prediksi menggunakan model yang ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model (Utami, 2017). Pada penelitian ini proses menentukan bagian kalimat sebagai anggota kelas positif atau kelas negatif berdasarkan nilai perhitungan Support Vector Machine dan K-Nearest Neighbor. Jika hasilnya kalimat tersebut untuk kelas positif lebih besar dari pada kelas negatif, maka kalimat tersebut termasuk ke dalam kelas positif begitu sebaliknya. 1. Algoritma Support Vector Machine (SVM) Nilai training cycles dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai training cycles. Hasil pengujian menunjukkan bahwa penerapan metode Support Vector Machine pada Tabel 7 dengan C = 0.0 dan Epsilon E = 0.0 dihasilkan Accuracy= 88.00% dan AUC= 0.964. Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan tweet/ komentar yang menggunakan tagar #2019GantiPresiden menggunakan metode SVM pada framework RapidMiner. 2. Algoritma SVM Berbasis PSO Hasil training cycles menggunakan metode SVM berbasis PSO dapat dilihat pada tabel 9 dibawah ini. Hasil terbaik pada eksperimen SVM berbasis PSO diatas adalah C=0.9 dan Epsilon E=0.9 serta population size=5 menghasilkan accuracy=92.75% dan AUC=0.973. Hal ini menunjukkan bahwa dengan menggunakan optimasi Particle Swarm Optimization dapat meningkatkan akurasi yang lebih baik. Nilai AUC dan akurasi yang dihasilkan setelah dilakukan optimasi menggunakan algoritma PSO meningkat, seperti yang terlihat pada tabel.9 dan gambar.3 dibawah ini. 3. Algoritma K-Nearest Neighbor (k-NN) Nilai k yang digunakan menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada data uji. Untuk memperkirakan nilai k yang terbaik, bisa dilakukan dengan menggunakan teknik validasi silang (Cross Validation). Hasil pengujian menunjukkan bahwa penerapan metode k-Nearest Neighbor pada tabel 11 dengan penentuan nilai k=10 menghasilkan Accuracy= 88.50% dan AUC= 0.948 adalah nilai yang paling tertinggi. Kurva ROC tersebut diartikan dengan False Positive untuk garis horizontal dan True Positive untuk garis vertikal dengan nilai AUC= 0.948. 4. Algoritma k-NN Berbasis PSO Hasil pengujian dengan menggunakan metode k-NN berbasis PSO dapat dilihat pada tabel 13 dibawah ini. KESIMPULAN Berdasarkan pengujian yang telah dilakukan terhadap tweet fenomena tagar #2019GantiPresiden yang dikumpulkan menggunakan Rapidminer Studio sebanyak 400 data (200 positif dan 200 negatif). Dengan menggunakan metode Support Vector Machine (SVM), Support Vector Machine berbasis Particle Swarm Optimization (SVM+PSO), k-Nearest Neighbor (k-NN), dan k-Nearest Neighbor berbasis Particle Swarm Optimization (k-NN+PSO) maka hasilnya adalah hipotesa awal tidak sesuai dengan hasil akhir. Dalam penelitian ini, hasil perhitungan metode SVM memiliki Accuracy sebesar 88.00% dan AUC sebesar 0.964 sedangkan Metode SVM+PSO menghasilkan Accuracy sebesar 92.75% dan AUC sebesar 0.973. Pengujian juga telah

0.904 sedangkan metode SVM+PSO menghasilkan Accuracy sebesar 92.75% dan AUC sebesar 0.975. Pengujian juga telah dibandingkan dengan menggunakan metode k-NN dan k-NN berbasis PSO. Hasil perhitungan yang diperoleh dari pengujian data dengan metode k-NN yaitu menghasilkan Accuracy sebesar 88.50% dan AUC sebesar 0.948. Sementara itu metode k-NN berbasis PSO menghasilkan nilai Accuracy yang justru menurun yaitu sebesar 75.25% dan AUC sebesar 0.768. Penerapan Particle Swarm Optimization (PSO) terbukti dapat meningkatkan akurasi algoritma SVM pada klasifikasi tweet fenomena tagar #2019GantiPresiden untuk mengidentifikasi antara tweet positif dan tweet negatif. Sedangkan untuk algoritma k-NN justru menghasilkan akurasi yang sebaliknya yaitu lebih rendah dibandingkan algoritma k-NN tanpa PSO. Penerapan optimasi PSO pada algoritma k-NN mengalami penurunan sebesar 13%. Hal ini dapat disimpulkan bahwa dalam penelitian text mining penerapan optimasi menggunakan PSO belum tentu dapat mengoptimalkan nilai akurasi terutama pada algoritma k-NN. Mengingat banyaknya penelitian terdahulu yang telah menerapkan text mining berbahasa Inggris dengan sukses meningkatkan nilai akurasi k-NN menggunakan PSO, maka dapat dianalisa bahwa optimasi PSO pada algoritma k-NN dengan menggunakan dataset Bahasa Indonesia tidak dapat meningkatkan nilai akurasi. Metode SVM terbukti lebih unggul dalam klasifikasi teks tweet fenomena #2019GantiPresiden ini karena SVM bekerja dengan mencari parameter hyperplane yang terbaik yaitu nilai C dan Epsilon sehingga ada banyak kemungkinan akurasi dapat lebih optimal. Namun waktu pengujian data lebih lama dilakukan oleh SVM+PSO dibandingkan metode KNN+PSO. Saran kedepan untuk penelitian selanjutnya dapat dilakukan dengan menggunakan dataset dari domain yang berbeda, misalnya review produk, review film, review saham dan lain sebagainya dengan teks berbahasa Indonesia dan meneliti optimasi perbaikan dari KNN+PSO untuk meningkatkan akurasi serta menggunakan teknik n-grams dengan tiga term (three-grams) pada tahap preprocessing untuk meningkatkan performa klasifikasi sebagai kontribusi penelitian lanjutan. Menggunakan teknik pengklasifikasian teks yang lain dan metode pemilihan fitur yang lainnya, seperti Chi Square, Gini Index, Mutual Information, dan lain-lain agar hasil akurasinya dapat dibandingkan.

| Sources | Similarity |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <p>Mengenal istilah text miningCompare text</p> <p>text mining dapat menganalisis dokumen, mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta menentukan kesamaan di antara dokumen untuk mengetahui bagaimana mereka berhubungan dengan variabel lainnya (statsoft, 2015).</p> <p>http://textmining-center.blogspot.com/2016/07/mengenal-istilah-text-mining.html</p> | 10% |
| <p>Rancang Bangun Sistem Analisis Sentimen Data Twitter Berbahasa...Compare text</p> <p>Analisis Sentimen atau biasa disebut Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, clustering, information extraction dan information retrieval (Berry & Kogan, 2010). Dalam penelitian ini penulis menerapkan proses text mining dan proses...</p> <p>https://eprints.uns.ac.id/36990/</p> | 10% |