

Comparison of the Performance of Machine Learning Algorithms for Sarcasm Detection in Bahasa

Perbandingan Kinerja Algoritma Machine Learning Untuk Mendeteksi Kalimat Sarkasme Dalam Bahasa Indonesia

Mochamad Alfian Rosid^{1*}, Fajar Muharram², Ghozali Rusyid Affandi³

¹Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo, Indonesia, ²Fakultas Bisnis Hukum Dan Sosial, Universitas Muhammadiyah Sidoarjo, Indonesia, ³Fakultas Psikologi Dan Ilmu Pendidikan, Universitas Muhammadiyah Sidoarjo, Indonesia

Abstract. *Twitter has become a widely used social media. The amount of data held has led to research such as sentiment analysis. Sentiment analysis has a problem when there are sarcasm sentences, the polarity of the sentiment that should be negative, becomes positive sentiment due to the use of sarcasm sentences. The purpose of this study is to compare the performance of three machine learning methods, namely Support Vector Machine, Random Forest, and K-Nearest Neighbor to detect sarcasm sentences on Twitter social media. These three methods were chosen because they have a good performance in text classification. The dataset used is taken from Indonesian language twitter with crawling technique. From the results of the study, it was found that the Support Vector Machine method had the best performance with a recall value of 0.97, precision 0.98 and f1-score 0.98.*

Keywords: *Twitter, Support Vector Machine, Random Forest, K-Nearest Neighbor*

Abstrak. Twitter menjadi media sosial yang banyak digunakan. Besarnya data yang dimiliki, memunculkan penelitian-penelitian seperti analisis sentimen. Analisis sentimen memiliki permasalahan ketika terdapat kalimat sarkasme, polaritas sentimen yang seharusnya negatif, menjadi sentimen positif akibat penggunaan kalimat sarkasme. Tujuan dari penelitian ini membandingkan kinerja tiga metode machine learning, yaitu Support Vector Machine, Random Forest, dan K-Nearest Neighbor untuk mendeteksi kalimat sarkasme pada media sosial twitter. Ketiga metode ini dipilih karena memiliki performa yang cukup bagus dalam klasifikasi teks. Dataset yang digunakan diambil dari twitter berbahasa Indonesia dengan teknik crawling. Dari hasil penelitian didapatkan metode Support Vector Machine memiliki kinerja terbaik dengan nilai recall 0.97, precision 0.98 dan f1-score 0.98.

Kata kunci: Twitter, Support Vector Machine, Random Forest, K-Nearest Neighbor

Pendahuluan

Analisis sentimen memiliki banyak kesulitan dalam penerapannya, salah satu diantaranya adalah mendeteksi sarkasme dari suatu teks. Sarkasme adalah jenis sentimen di mana orang menyampaikan pesan mereka secara implisit. Berisi perselisihan positif untuk menyampaikan pendapat negatif [1]. Dalam kalimat sarkasme, hal yang terjadi justru berbeda dimana kalimat sarkasme umumnya justru bermakna positif tetapi menggambarkan situasi yang sebaliknya [2]. Kegagalan mengidentifikasi sarkasme di dalam teks dapat mempengaruhi hasil pengklasifikasian sentiment analysis dari suatu dataset. Berdasarkan hasil penelitian yang dilakukan oleh Antonakaki dkk.[3] pada saat masa pemilihan presiden Amerika Serikat, tercatat 11% pengguna Twitter yang aktif dalam isu pemilihan presiden tersebut menggunakan kalimat sarkasme dalam mengutarakan pendapatnya. Hal ini membuktikan bahwa penggunaan majas sarkasme pada sosial media Twitter masih sering ditemukan.

Sejauh ini telah ada beberapa penelitian terdahulu terkait masalah deteksi sarkasme menggunakan pendekatan machine learning dan deep learning dalam analisis sentiment. [4] menyajikan metode deteksi sarkasme untuk analisis sentimen tweet berbahasa Indonesia. Deteksi sarkasme menggunakan metode Random Forest. Penelitian lain menggunakan One Against One(OAO) dan One Against All (OAA) [5], hasilnya SVM OAA memiliki kinerja lebih baik dibandingkan dengan SVM OAO. Sedangkan [6] menyajikan metode untuk mendeteksi sarkasme menggunakan deep learning pada tweet kombinasi bahasa Inggris dengan bahasa asli

India, Hindi. Model yang diusulkan adalah gabungan dari bidirectional long short-term memory dengan lapisan perhatian softmax dan convolution neural network untuk deteksi sarkasme.

Penelitian [7] mengusulkan kombinasi pengklasifikasi machine learning dan deep learning untuk mengekstrak target sarkasme dari text. Dalam penelitian ini menggunakan tiga kumpulan data yang tersedia secara umum yaitu cuplikan buku sarkasik, tweet sarkastik dan komentar reddit. Pada pendekatan machine learning digunakan untuk mengklasifikasi kalimat sarkastik dan menentukan apakah sebuah kalimat sarkastik mengandung target. Kemudian menggunakan model deep learning dari Analisis Sentimen Berbasis Aspek untuk mengekstrak target.

Dari penelitian-penelitian yang disebutkan diatas baik menggunakan metode machine learning dan deep learning mampu melakukan deteksi sarkasme dengan cukup baik. Pada penelitian ini mencoba untuk melakukan perbandingan kinerja dari metode machine learning. Metode yang akan dibandingkan adalah Random Forest (RF), Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Diharapkan dari hasil uji coba, metode machine learning dengan kinerja terbaik dapat digunakan sebagai referensi dalam pembuatan model deteksi sarkasme pada media sosial twitter.

Makalah ini disusun sebagai berikut. Pada bagian 1 menunjukkan pendahuluan. Bagian 2 membahas metode machine learning untuk deteksi sarkasme. Bagian 3 merupakan metode penelitian. Bagian 4 mengungkapkan hasil dan pembahasan dari penelitian dan analisis perbandingan kinerja. Bagian 5 adalah tentang kesimpulan dan pekerjaan masa depan dari penelitian ini.

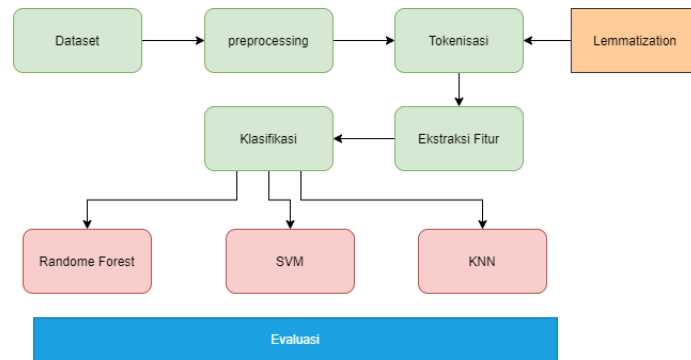
Deteksi Sarkasme Berdasarkan Metode Machine Learning

Pada metode machine learning terdapat tiga konsep, yaitu *Supervised Learning*, *weakly Supervised Learning*, dan *Unsupervised Learning*. *Supervised Learning* adalah model pembelajaran berdasarkan label atau target yang sudah ditentukan, evaluasi model berdasarkan perbandingan hasil prediksi dengan target. *Weakly Supervised Learning* merupakan model pembelajaran dimana model dilatih dengan contoh yang hanya sebagian diberi label. *Unsupervised Learning* berbeda dengan kedua konsep tersebut, *Unsupervised Learning* tidak dilatih dengan data yang telah diberi label. Pada penelitian ini mengadopsi metode RF, SVM dan KNN.

RF merupakan model *ensemble* yang menggunakan bagging untuk pengambilan sample, sehingga kinerjanya sangat bagus [8], SVM dikenal sebagai pengklasifikasi terbaik yang memberikan hasil paling akurat dalam masalah klasifikasi ungkapan. SVM mencapainya dengan membuat hyperplane dengan jarak Euclidean maksimal untuk contoh terlatih terdekat. Hyperplane Support Vector Machine sepenuhnya diselesaikan oleh subset yang relatif kecil dari data training yang diperlakukan sebagai vektor pendukung. Dataset training yang tersisa tidak memiliki akses ke pengklasifikasi yang memenuhi syarat. Jadi untuk tujuan klasifikasi teks, SVM pengklasifikasi telah berhasil diterapkan dan juga digunakan dalam aplikasi pemrosesan urutan yang berbeda. SVM digunakan dalam hiperteks dan klasifikasi teks karena tidak memerlukan dataset training berlabel [9]. KNN memiliki perhitungan yang sangat sederhana, karena kesederhanaannya, KNN banyak digunakan diberbagai bidang, namun KNN memiliki kekurangan apabila menangani data yang sangat besar [10].

Metode Penelitian

Tahapan penelitian diawali dari pengambilan data, kemudian dilakukan proses preprocessing, tokenisasi, pembobotan fitur, dan proses klasifikasi dengan membandingkan 3 (tiga) *classifier*. Gambar 1. Menunjukkan arsitektur sistem untuk metode machine learning.



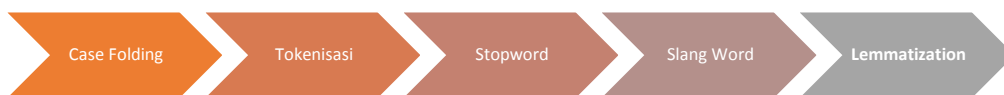
Gambar 1. Arsitektur Sistem untuk Metode Machine Learning

Dataset

Dataset diambil dari media sosial twitter dengan memanfaatkan API twitter, yang memudahkan para peneliti untuk pengambilan data twitter. Data diambil dengan kata kunci #Anis Baswedan, #Cebong, #Jokowi. Dimana tweet yang mengandung kata kunci tersebut sering menggunakan kalimat sarkasme.

Preprocessing

Pada penelitian ini, proses preprocessing menggunakan tahapan standar dalam klasifikasi teks. Preprocessing dimulai dari *case folding*, Tokenisasi, *Stopword*, *Slang Word*, dan *Lemmatization*. Gambar 2 menunjukkan proses preprocessing. Dari hasil preprocessing, dapat ditampilkan word cloud kata yang sering muncul seperti yang diperlihatkan oleh gambar 3.



Gambar 2. Proses Preprocessing

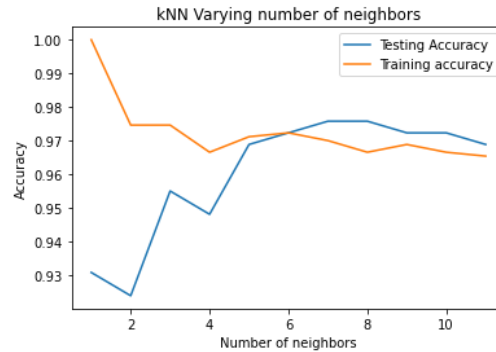


Gambar 3. Word Cloud Kata Yang Sering Muncul

Hasil dan Pembahasan

Pada tahapan ini, ketiga metode dibandingkan kinerjanya, diawali dengan mencoba melakukan pengujian terhadap data menggunakan metode Randome Forest, kemudian dilanjutkan melakukan pengujian dengan metode SVM dan terakhir data akan diuji menggunakan metode KNN. Model yang telah dibangun dilakukan proses pengujian dan evaluasi antara data training dengan data testing. Kemudian pada tahap evaluasi, penelitian ini menggunakan pengukuran Accuracy, precision, recall, dan f1-score. Untuk pengujian menggunakan metode KNN, pada penelitian ini mencoba mencari nilai k (jumlah tetangga) terbaik antara 1-10. Gambar 4 merupakan hasil ujicoba menggunakan KNN dengan variasi jumlah tetangga antara 1-10. KNN dengan jumlah tetangga k=7 memiliki nilai akurasi terbaik yaitu 97.57%.

Setelah semua pengujian selesai, maka didapat hasil akurasi untuk masing-masing metode yang digunakan. Tabel 1. Menunjukkan hasil perbandingan kinerja dari masing-masing metode. Berdasarkan Tabel 1 didapat nilai akurasi tertinggi pada metode SVM dengan nilai akurasi 0.98, precision 0.98, recall 0.97 dan f1-score 0.98. Hasil dari metode SVM tidak jauh berbeda dengan yang dihasilkan oleh metode random forest dengan nilai akurasi 0.97, precision 0.97, recall 0.96 dan f1-score 0.97. Hasil terendah didapat oleh metode KNN dengan nilai akurasi adalah 0.95, nilai precision 0.96, recall 0.90 dan f1-score 0.93.



Gambar 4. Hasil KNN dengan Variasi Jumlah Tetangga

Tabel 1. Hasil Perbandingan Kinerja

Metode	Accuracy	Precision	Recall	F1-Score
Randome Forest	0.97	0.97	0.96	0.97
SVM	0.98	0.98	0.97	0.98
KNN	0.95	0.96	0.90	0.93

Kesimpulan

Berdasarkan hasil pengujian yang telah dibahas pada bagian hasil dan pembahasan, maka diperoleh metode terbaik untuk deteksi sarkasme adalah metode klasifikasi SVM, hal ini dikarenakan ketepatan klasifikasinya yang paling tinggi daripada yang lain dengan nilai akurasi 0,98 atau 98%, nilai precision 0,98, nilai recall 0,97 dan nilai f1 adalah 0,98. Pada pekerjaan masa depan, akan diteliti kinerja hasil penggabungan ketiga metode machine learning dengan teknik Ensemble Stacking Classifier.

Ucapan terima kasih

Dengan ini kami ucapkan terima kasih kepada Universitas Muhammadiyah Sidoarjo yang telah membiayai dan mendukung publikasi penelitian ini.

Referensi

- [1] K. Rajeswari and P. S. Bala, "Recognition of Sarcastic Emotions of Individuals on Social Network," *Int. J. Pure Appl. Math.*, vol. 118, no. 7 Special Issue, pp. 253–258, 2018.
- [2] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation," *EMNLP 2013 - 2013 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. October, pp. 704–714, 2013.
- [3] D. Antonakaki, D. Spiliotopoulos, C. V. Samaras, P. Pratikakis, S. Ioannidis, and P. Fragopoulou, "Social media analysis during political turbulence," *PLoS One*, vol. 12, no. 10, pp. 1–23, 2017.
- [4] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 1, p. 53, 2019.
- [5] D. Alita, "Multiclass SVM Algorithm for Sarcasm Text in Twitter," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 118–128, 2021.
- [6] D. Jain, A. Kumar, and G. Garg, "Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN," *Appl. Soft Comput. J.*, vol. 91, p. 106198, 2020.
- [7] P. Parameswaran, A. Trotman, V. Liesaputra, and D. Evers, "Detecting the target of sarcasm is hard: Really?," *Inf. Process. Manag.*, vol. 58, no. 4, p. 102599, 2021.
- [8] V. K. Gupta, A. Gupta, D. Kumar, and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," *Big Data Min. Anal.*, vol. 4, no. 2, pp. 116–123, 2021.
- [9] B. K. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis," *Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017*, no. Icicct, pp. 216–221, 2017.
- [10] W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," *IEEE Access*, vol. 8, pp. 28808–28819, 2020.