

SENTIMENT ANALYSIS PADA REVIEW PENGGUNA APLIKASI SNAPCHAT DENGAN VADER DAN ALGORITMA MACHINE LEARNING

Sentiment Analysis on User Reviews Of The Snapchat Application with Vader and Machine Learning Algorithms

Padel Mohammad Agam¹⁾, Muhammad Idzha Adhitya Ranius²⁾, Tata Sutabri³⁾ A. Yani Ranius⁴⁾

^{1,2,3,4}Program Studi Magister Teknik Informatika, Pascasarjana, Universitas Bina Darma, Universitas Bina Darma, Jl. Jenderal Ahmad Yani No. 12 Plaju, Palembang, Sumatera Selatan

*email: padelmohammad.agam@gmail.com¹⁾, idzhaadhityaranius@gmail.com²⁾
tata.sutabri@binadarma.ac.id³⁾, ay_ranius@binadarma.ac.id⁴⁾

Abstract

The development of technology in almost every aspect of life, one of which is the social aspect which is currently heavily influenced by the development of digital technology in the form of social media to interact and communicate with other people online. There are several social media that are used by the younger generation today, one of which is social media Snapchat which is a social media application used to share photos and videos to a list of recipients by sending temporary messages and stories that can be viewed for a limited time as well as unique features. others are still growing. The development of a number of features on Snapchat certainly requires a review from him with various reviews and feedback from users so it is necessary to do sentiment analysis on Snapchat application user reviews using the VADER approach and machine learning algorithms. In the early stages of the research, the review data collection was in the form of a dataset review of snapchat application users to data labeling which was then carried out in the data training and data testing stages using the SVM (Support Vector Machines) algorithm so that the research results could identify positive, negative, or neutral sentiments from user reviews. in assisting app developers in improving user experience and addressing issues that may arise with the Snapchat app.

Keywords : Machine Learning, Sentiment Analysis, Vader, Snapchat

Abstrak

Perkembangan teknologi hampir di setiap aspek kehidupan, salah satunya pada aspek sosial yang saat ini sangat dipengaruhi dengan perkembangan teknologi digital berupa media sosial untuk berinteraksi dan berkomunikasi dengan orang lain secara online. Terdapat sejumlah media sosial yang digunakan oleh generasi muda saat ini salah satunya media sosial Snapchat yang merupakan sebuah aplikasi media sosial yang digunakan untuk berbagi foto dan video ke daftar penerima dengan pengiriman pesan yang sementara waktu dan cerita yang dapat dilihat dalam jangka waktu terbatas serta fitur unik lainnya yang terus berkembang. Perkembangan sejumlah fitur-fitur pada Snapchat tentunya membutuhkan review dari penggunanya dengan beragam ulasan dan umpan balik dari pengguna sehingga perlu dilakukannya analisis sentimen pada review pengguna aplikasi Snapchat menggunakan pendekatan VADER dan algoritma machine learning. Pada tahap awal penelitian, kumpulan data review berupa Dataset review pengguna aplikasi snapchat hingga labelling data yang kemudian dilakukan tahapan data training dan data testing menggunakan algoritma SVM (*Support Vector Machines*) sehingga hasil penelitian dapat mengidentifikasi sentimen yang positif, negatif, atau netral dari review pengguna dalam membantu pengembang aplikasi dalam meningkatkan pengalaman pengguna dan mengatasi masalah yang mungkin muncul pada aplikasi Snapchat.

Kata kunci: Machine Learning, Sentiment Analysis, Vader, Snapchat

1. Pendahuluan

Beberapa dari kita mungkin ada yang belum mengetahui tentang aplikasi sosial media yang baru – baru ini sedang populer. Jika kita sudah pernah mengenal Instagram mungkin akan familiar dengan aplikasi ini. Namanya aplikasi Snapchat, fiturnya melebihi Instagram dalam beberapa bagian. Telah didownload lebih dari 100.000.000 pengguna android. Snapchat yang mulai populer tahun 2014 silam kini semakin bertambah penggunanya. Snapchat dulu bernama Picabbo di buat oleh 3 mahasiswa Stanford University yaitu Evan Spiegel, Bobby Murphy dan Reggie Brown pada tahun 2011.

Pada tahun 2012 Picabbo berubah nama menjadi Snapchat hingga sekarang. Selama itu, Snapchat telah mencatat 700 juta foto dan video di bagikan setiap hari. Aplikasi yang mempunyai logo hantu berlatar kuning ini mempunyai kemiripan dengan aplikasi Instagram. Mereka sama – sama sosial media yang mempunyai konsep berbagi foto dan video. Jika instagram bisa membagi foto atau video yang ada di gallery, maka snapchat membagi foto atau video langsung melalui kamera Android atau iPhone.

Setiap aplikasi yang terus bertumbuh membutuhkan review dari penggunanya, review bisa saja positif, negatif maupun netral. Seperti mengekspresikan sentimen positif atas suatu produk atau sentimen negatif untuk seorang tokoh politik. Pendekatan untuk mengekstraksi informasi opini yang terkandung dalam data text dari review pengguna aplikasi snapchat secara otomatis dengan pendekatan komputasional menjadi penting dilakukan [1].

Menurut Pang & Lee pendekatan untuk mengekstraksi informasi opini yang terkandung dalam suatu data text disebut sebagai Sentiment Analysis. Sentiment analysis termasuk bidang penelitian di bawah Natural Language Processing. Karena berkaitan dengan natural language, upaya untuk meningkatkan kinerja sentiment analysis bisa dengan mengekstraksi aspek kontekstual dalam data text. Dalam melakukan sentiment analysis dapat menggunakan Algoritma Support Vector Machine (SVM) yang merupakan salah satu metode untuk melakukan klasifikasi ataupun prediksi. Pada prinsipnya, metode ini akan mencari ruang pemisah yang paling optimal dalam suatu dataset dengan kelas yang berbeda [2].

Sebagian besar penelitian sentiment analysis fokus pada dataset product review sehingga fokus penelitian berkaitan dengan sentiment analysis pada review pengguna aplikasi Snapchat menjadi penting dilakukan untuk masa mendatang.

2. Tinjauan Pustaka

Sentiment analysis adalah proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut [3]. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber. Sentiment analysis berperan sebagai alat yang dapat menghubungkan seluruh data tersebut. Dengan begitu, perusahaan dapat memperoleh masukan inti dari pengguna atau konsumen secara efisien [4].

Sentiment analysis merupakan salah satu bidang dari Natural Language Processing (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi review [5].

Dengan bantuan sentiment analysis, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur. Data tersebut dapat menjelaskan opini masyarakat mengenai produk, merek, layanan, politik, atau topik lainnya. Perusahaan, pemerintah, maupun bidang lainnya kemudian memanfaatkan data-data tersebut untuk membuat analisis marketing, review produk, umpan-balik produk, dan layanan masyarakat [6].

Guna menghasilkan opini yang dibutuhkan, sentiment analysis tidak hanya harus bisa mengenali opini dari teks. Proses yang juga disebut sebagai opini mining ini juga perlu bekerja dengan mengenali tiga aspek berikut:

1. Subjek: topik apa yang sedang dibicarakan.
2. Polaritas: apakah opini yang diberikan bersifat positif atau negatif.
3. Pemegang opini: seseorang yang mengeluarkan opini tersebut.

Sentiment analysis kemudian akan membedakan teks menjadi dua kategori, yakni fakta dan opini. Fakta merupakan ekspresi objektif mengenai sesuatu. Sementara opini adalah ekspresi subjektif yang menggambarkan sentimen, perasaan, maupun penghargaan terhadap suatu hal [7].

Dalam tulisannya, C.J. Hutto dan Eric Gilbert (2014) mengatakan bahwa Vader Lexicon Polarity Detection adalah sebuah model sederhana dengan berlandaskan rule atau aturan untuk analisa sentimen secara umum[8].

Metode leksikal tidak memerlukan data training ataupun data yang telah diberikan label, namun sudah tersedia kamus lengkap yang berisi polaritas sentimen sehingga ia dapat menentukan sentimen data dengan cepat [9].

Dalam penelitian Gilbert dan Hutto (2014) juga memperlihatkan hasil analisa sentimen dari VADER mendapat skor yang tinggi dalam kategori media sosial, namun tidak cukup baik dalam kategori review product, dalam penelitian kami, model ini digunakan untuk memberikan analisa sentimen terhadap review pengguna aplikasi snapchat [10].

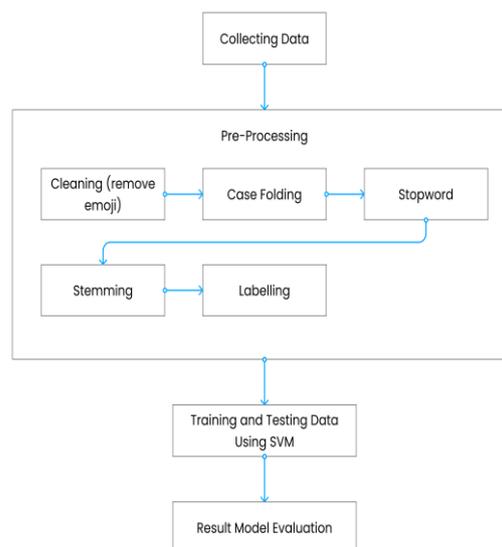
Algoritma Random Forest adalah salah satu metode berbasis klasifikasi dan regresi dimana terdapat proses agregasi pohon keputusan. Random Forest bekerja dengan membangun beberapa decision tree dan menggabungkannya demi mendapatkan prediksi yang lebih stabil dan akurat. ‘Hutan’ yang dibangun oleh Random Forest adalah kumpulan decision tree di mana biasanya dilatih dengan metode bagging [11]. Ide umum dari metode bagging adalah kombinasi model pembelajaran untuk meningkatkan hasil keseluruhan.

Algoritma Support Vector Machine (SVM) merupakan salah satu metode untuk melakukan klasifikasi ataupun prediksi. Pada prinsipnya, metode ini akan mencari ruang pemisah yang paling optimal dalam suatu dataset dengan kelas yang berbeda. Dengan memaksimalkan jarak antar kelas, metode SVM akan mencari hyperplane yang terbaik [12].

Dimana Hyperplane yang dimaksud merupakan suatu fungsi yang dapat digunakan sebagai pemisah antar kelas. Dalam dua dimensi fungsi yang digunakan untuk klasifikasi antar kelas disebut sebagai line whereas, fungsi yang digunakan untuk klasifikasi antar kelas dalam tiga dimensi disebut plane similarly, sedangkan fungsi yang digunakan untuk klasifikasi di dalam ruang kelas dimensi yang lebih tinggi disebut hyperplane [13]. Dalam SVM objek data terluar yang paling dekat dengan hyperplane disebut support vector, objek inilah yang paling sulit diklasifikasikan karena posisi yang hampir tumpang tindih dengan kelas lain [14].

3. Metodologi Penelitian

Metode yang digunakan dalam analisis sentiment secara garis besar melalui beberapa tahapan yang digambarkan dalam bentuk alur kerja pada Gambar 1 yang tersusun atas sejumlah tahapan yang perlu dilakukan meliputi pengambilan data, tahap pre-processing data berupa data *cleaning*, *case folding*, *filter stopword*, *stemming*, dan *labelling*, yang selanjutnya diperoleh sentimen dari semua data yang berisi kata paling populer yang muncul pada review aplikasi snapchat, kemudian dataset selanjutnya akan dilakukan tahapan data training dan data testing menggunakan algoritma SVM untuk dilihat hasil evaluasi data.



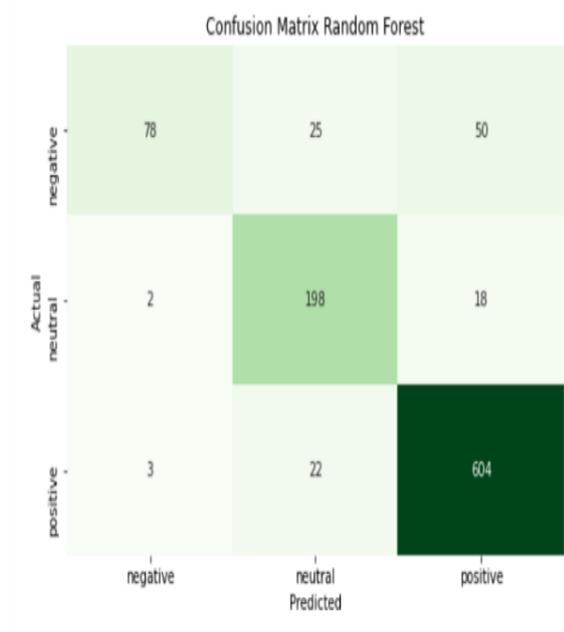
Gambar 1. Tahapan Penelitian

4. Hasil dan Pembahasan

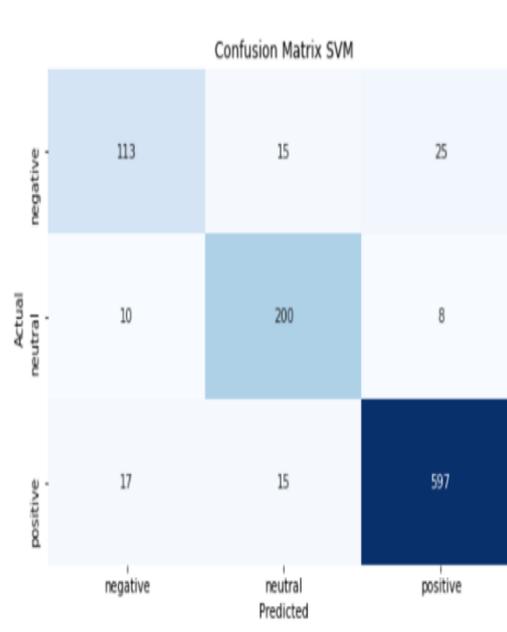
Akumulasi dataset review pengguna aplikasi Snapchat pada google play store diambil dari situs Kaggle, dimana data tersebut disediakan oleh Umar Pervaiz, dengan total 32859 baris data review. Pada tahapan selanjutnya dilakukan Pre-processing, dataset dibatasi sebanyak 5000 baris data dan kolom yang tidak diperlukan juga dihapus. Selanjutnya data dirapikan dengan mengubah semua jenis huruf menjadi huruf kecil (*case folding*), lalu dilanjutkan dengan tahapan menghilangkan *stopword* dan *stemming*, yaitu mengubah kata yang ada menjadi kata dasar. Langkah terakhir pada tahapan ini yaitu menentukan skor masing-masing data lalu menentukan label sentimen setiap data tersebut menggunakan VADER (Valence Aware Dictionary And Sentiment Reasoner).

Setelah melalui tahapan pre-processing data, berikutnya dilakukan tahapan klasifikasi terhadap analisis sentiment yang sebelumnya telah didapat. Pada tahap ini diterapkan algoritma Random Forest dan SVM. Setelah proses klasifikasi analisis sentimen telah selesai, langkah selanjutnya adalah mengevaluasi dengan mengukur tingkat akurasi dan kualitas dari hasil tersebut. Evaluasi yang dilakukan adalah dengan pengujian performa dan akurasi sehingga menghasilkan nilai *accuracy*, *precision*, dan *recall* dalam bentuk *confusion matrix*.

Hasil pelabelan sentimen dengan menggunakan VADER terhadap dataset review pengguna aplikasi snapchat pada google play store menghasilkan sentimen Positive sebanyak 3092, sentimen Neutral sebanyak 1039, dan sentimen Negative sebanyak 869 data.



Gambar 6. Confusion Matrix Random Forest



Gambar 8. Confusion Matrix SVM

Dataset kemudian berikutnya dijalankan dengan algoritma Support Vector Machine, hasil yang didapatkan adalah akurasi sebesar 91% dengan rincian sebagai berikut:

	precision	recall	f1-score	support
negative	0.81	0.74	0.77	153.00
neutral	0.87	0.92	0.89	218.00
positive	0.95	0.95	0.95	629.00
accuracy	0.91	0.91	0.91	0.91
macro avg	0.87	0.87	0.87	1000.00
weighted avg	0.91	0.91	0.91	1000.00

Gambar 7. SVM

Confusion Matrix memperlihatkan terdapat 597 True Positive yang didapatkan dengan algoritma SVM, sementara terdapat 200 True Neutral dan 113 True Negative.

5. Kesimpulan

Adapun kesimpulan yang dapat ditarik dalam penelitian ini didasarkan pada pelabelan sentimen menggunakan VADER, dengan dataset yang berjudul Review Pengguna Aplikasi Snapchat pada Google Play Store. Dataset sentimen tersebut kemudian diklasifikasi menggunakan 2 algoritma machine learning, yaitu Random Forest dan SVM, dimana algoritma SVM memberikan hasil yang lebih baik dengan nilai akurasi 91%. Sementara algoritma Random Forest menghasilkan nilai akurasi sebesar 88%. Dengan demikian penelitian ini dapat menemukan hasil yang sesuai dengan yang diinginkan sesuai kebutuhan pengguna dari Aplikasi Snapchat pada Google Play Store dengan menggunakan VENDER dan Algoritma Machine Learning.

Daftar Rujukan

- [1] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram," *J-SAKTI (Jurnal Sains Komput. dan Inform., vol. 2, no. 2, p. 200, 2018, doi: 10.30645/j-sakti.v2i2.83.*
- [2] A. A. Prasanti, M. A. Fauzi, and M. T. Furqon, "Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode N- Gram dan Neighbor Weighted K-Nearest Neighbor (NW-

- KNN),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. Vol. 2, no. 2, pp. 594–601, 2018.
- [3] R. Maipradit, H. Hata, and K. Matsumoto, “Sentiment Classification Using N-Gram Inverse Document Frequency and Automated Machine Learning,” *IEEE Softw.*, vol. 36, no. 5, pp. 65–70, 2019, doi: 10.1109/MS.2019.2919573.
- [4] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes,” *2018 Int. Conf. Orange Technol. ICOT 2018*, no. October, 2018, doi: 10.1109/ICOT.2018.8705796.
- [5] S. H. Badjrie, O. N. Pratiwi, and H. D. Anggana, “Analisis Sentimen Review Customer Terhadap Produk Indihome Dan First Media Menggunakan Algoritma Convolutional Neural Network Review Analysis Sentiment Customer Product Indihome And First Media Using Convolutional Neural Network,” *eProceedings ...*, vol. 8, no. 5, pp. 9049–9061, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversit y.ac.id/index.php/engineering/article/view/15877>
- [6] S. Vashishtha and S. Susan, “Fuzzy rule based unsupervised sentiment analysis from social media posts,” *Expert Syst. Appl.*, vol. 138, no. July, 2019, doi: 10.1016/j.eswa.2019.112834.
- [7] B. Xiang, C. Zeng, X. Dong, and J. Wang, “The application of a decision tree and stochastic forest model in summer precipitation prediction in Chongqing,” *Atmosphere (Basel)*, vol. 11, no. 5, 2020, doi: 10.3390/ATMOS11050508.
- [8] F. Azhiman, R. N. Dasmien, A. Putra, and W. Agustian, “Collecting Data Desa di Kecamatan Rambutan Kabupaten Banyuasin untuk Pengimplementasian Sistem Digital Desa,” vol. 2, no. 6, pp. 1733–1742, 2022.
- [9] S. Elbagir and J. Yang, “Language Toolkit and VADER Sentiment,” *Proc. Int. MultiConference Eng. Comput. Sci.*, vol. 0958, pp. 12–16, 2019.
- [10] S. Sunarto *et al.*, “ADOPSI WEB 4 . 0 UNTUK CITIZEN LEARNING DALAM PENERAPAN E-GOVERNMENT KOTA PALEMBANG WEB 4 . 0 ADOPTION FOR CITIZEN LEARNING IN APPLICATION E-GOVERNMENT OF PALEMBANG,” vol. 0, no. 02, pp. 42–47, 2022.
- [11] F. Khasan *et al.*, “Pengukuran Usability pada Website SMK Ma ’ arif 3 Somalangu Kebumen Menggunakan System Usability Scale (SUS) Usability Measurement on the Website of SMK Ma ’ arif 3 Somalangu Kebumen Using the System Usability Scale (SUS),” vol. 0, no. 02, pp. 67–72, 2022.
- [12] R. Yanto, Alfiarini, Veradilla Amalia, “Analisa Penentuan Reward Berbasis Kinerja Dosen Menggunakan Metode Simple Multi Atribute Rating Technique,” *J. Ilm. Bin. STMIK Bina Nusant. Jaya Lubuklinggau*, vol. 4, no. 2, pp. 80–85, 2022, doi: 10.52303/jb.v4i2.82.
- [13] F. Azhiman and R. N. Dasmien, “Sosialisasi Performansi Antena Omnidirectional Palembang PT . LEN REKAPRIMA SEMESTA LRT Kota,” vol. 5, no. 2, pp. 359–371, 2022.

- [14] A. Yasir, “Penerapan Metode Simple Additive Weighting dalam Penentuan Media Pembelajaran Online di Masa Pandemi Covid-19,” *J. Ilm. Bin. STMIK Bina Nusant. Jaya Lubuklinggau*, vol. 4, no. 2, pp. 94–98, 2022, doi: 10.52303/jb.v4i2.83.