

ANALISIS SENTIMEN TERHADAP LAYANAN APLIKASI DANA MENGUNAKAN NAÏVE BAYES PADA SOSIAL MEDIA TWITTER

Amanah Cetagati¹⁾, Rido Febryansyah²⁾

¹ Sistem Informasi Akutansi, ² Sistem Informasi

*) ridofebryansyah5@gmail.com

Abstrak

Seiring berkembangnya zaman, saat ini teknologi informasi berkembang semakin cepat dalam setiap bidang salah satunya dalam bidang keuangan. Kini transaksi pun sudah tidak dilakukan secara manual, mulai dari adanya mesin ATM, mobile banking sampai yang trend saat ini adalah e-wallet. Bertransaksi menggunakan e-wallet sangatlah mudah, kita hanya perlu bertukar nomor handphone yang sudah kita daftarkan pada aplikasi e-wallet tersebut. Aplikasi DANA merupakan salah satu aplikasi e-wallet yang paling diminati karena tampilan yang bagus dan cara menggunakan yang mudah. Akan tetapi walaupun diminati persepsi pengguna terdapat opini yang positif dan negatif. Biasanya pengguna memilih beropini secara terbuka melalui media sosial seperti twitter karena merasa lebih terbuka daripada berpendapat langsung. Untuk mengetahui berhasil atau tidaknya aplikasi DANA, maka diadakan penelitian analisis sentiment terhadap aplikasi dana di media sosial twitter menggunakan metode naïve bayes.

Kata Kunci: Analisis Sentimen, DANA, *e-wallet*, Twitter

PENDAHULUAN

Uang elektronik adalah alat pembayaran yang saat ini marak digunakan karena fungsinya yang fleksibel (Anggraini & Suaidah, 2022). Karena saat menggunakan uang elektronik kita tidak perlu membawa uang cash, transaksi menggunakan uang elektronik ini juga sangat mudah, saat kita akan menggunakannya, kita dapat hanya menginputkan nomor atau id si penerima (Anggriani & Husna, 2019);(Darwis et al., 2022);(Windane & Lathifah, 2021). Karena maraknya penggunaan uang elektronik untuk mencegah terjadinya kejahatan maka dikembangkan lah aplikasi dompet elektronik atau yang biasa disebut e-wallet (Lina et al., 2021);(Lina & Nani, 2020);(Middleton et al., 2000). E-wallet ini berfungsi sebagai tempat penyimpanan uang elektronik yang berbentuk aplikasi (Dheara et al., 2022). Pada e-wallet ini terdapat juga fitur-fitur seperti transfer, pembayaran, pembelian pulsa, token listrik dan lainnya yang sangat membantu memenuhi kebutuhan sehari-hari (Anestiviya et al., 2021).

Karena kebutuhan masyarakat yang cukup tinggi dalam penggunaan e-wallet berbagai macam jenis e-wallet diciptakan dengan fitur-fitur unggulan tertentu (Alita, 2021a). Salah

satu aplikasi e-wallet yang banyak digunakan di Indonesia adalah aplikasi DANA. Sejak diluncurkan bulan Desember tahun 2018 lalu, aplikasi DANA telah di download sebanyak 10 juta kali yang menandakan masyarakat Indonesia sangat antusias menggunakan aplikasi e-wallet ini dan dinilai sangat efektif. Untuk mengetahui respon dan komentar pengguna, DANA juga membuat media sosial sebagai wadah interaksi bagi pengguna-penggunanya untuk mengetahui bagaimana pelayanan aplikasi DANA dari sentimen pengguna, jadi pengguna aplikasi DANA dapat memberikan komentar atau keluhan di media sosial official DANA. Salah satu media sosial yang digunakan adalah media sosial twitter (Septilia et al., 2020);(Ramadhan & Setiawan, 2019).

Twitter merupakan media sosial yang berdiri sejak 2009, dengan banyaknya pengguna aktif twitter dapat dikatakan sebagai media sosial terpopuler yang digunakan dari berbagai macam kalangan (Fikri et al., 2020);(Styawati & Mustofa, 2019). Pada twitter, pengguna dapat menuangkan pikirannya pada fitur tweet. Oleh karena itu, penelitian analisis sentimen ini menggunakan media sosial twitter karena dinilai lebih efektif dan mudah untuk mengambil data dari setiap tweet pengguna (Mittal & Patidar, 2019);(Alita et al., 2020);(Isnain et al., 2021).

Dalam melakukan analisis sentimen dibutuhkan metode teks mining, pada penelitian ini, metode yang digunakan adalah metode Naïve Bayes Classifier dengan algoritma _ (SVM) (Nurkholis et al., 2022);(Syah & Witanti, 2022). Tujuan penelitian analisis sentimen ini dilakukan untuk mengetahui bagaimana tanggapan pengguna aplikasi DANA terhadap layanan DANA pada sosial media twitter .

METODE

Metodologi penelitian yang digunakan yaitu metode Knowledge Discovery in Databases (KDD) yang memiliki 5 tahap yaitu selection data , preprocessing, transformation, data mining, dan evaluation:

Selection Data

Pada tahapan awal ini akan dilakukan pencarian dan pengambilan data tweet dari Twitter yang relevan dengan penelitian (Wahyono et al., 2021);(Giovani et al., 2020). Data yang diambil menggunakan teknik web scraping dengan bantuan package Twitter Scraper dengan

tools Anaconda Prompt. Data yang diambil dari Twitter yaitu tweet dari bulan januari 2019 sampai bulan januari 2020 dengan kata kunci @DANA_id. Untuk proses scraping pada Twitterscraper

Preprocessing

Setelah data tweet tersebut selesai diseleksi, tahapan selanjutnya yaitu tahapan preprocessing (Indrayuni, 2019);(Lestari et al., 2018);(Rahman Isnain et al., 2021). Tahapan preprocessing bertujuan untuk membersihkan data yang kotor. Pada tahapan ini URL, mention, hashtag, angka, simbol, dan tanda baca akan dibersihkan (Darwis et al., 2021);(Aldino, Darwis, et al., 2021a). Selanjutnya mengubah kata-kata yang disingkat dan mengubah kata-kata gaul (slang word) menjadi kata yang sebenarnya sesuai kaidah bahasa Indonesia. Kemudian menghapus kata-kata yang tidak penting (stopword) seperti “yang”, “di”, dll. Setelah data dibersihkan selanjutnya masuk ke tahapan terakhir preprocessing yaitu tweet tersebut akan diberikan label dengan menggunakan teknik skoring .

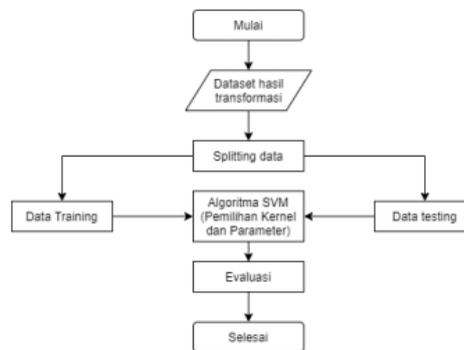
Transformation

Transformation merupakan tahap perubahan data, karena algoritma SVM hanya bisa memproses data numerik atau berbentuk vektor, maka data tweet yang sudah diberi label diubah menjadi data numerik dengan cara pembobotan kata (Abdul Maulud et al., 2021);(Prasetyawan et al., 2018). Tahapan term weighting atau pembobotan kata dilakukan dengan menghitung frekuensi kemunculan kata (term frequency) di dalam sebuah tweet dan kata di dalam tweet tersebut menjadi attribute (Styawati et al., 2021);(Sulistiani & Tjahyanto, 2016). Term Frequency (TF) merupakan salah satu skema pembobotan paling populer saat ini, 83% dari sistem rekomendasi berbasis teks di perpustakaan menggunakan TF. Untuk contoh perhitungan TF, Misalnya dataset yang akan diolah memiliki 2 tweet yaitu “cashback DANA baik baik” dan “pelayanan DANA mengecewakan”, dari data tersebut maka term frequency.

Data Mining

Tahapan data mining merupakan tahapan penerapan algoritma untuk klasifikasi (Herlinda et al., 2021);(Aldino, Darwis, et al., 2021b). Pada tahapan ini sebelum mulai melakukan

penerapan algoritma, kita harus membagi terlebih dahulu dataset menjadi data training dan data testing (Tuhuteru, 2020);(Alita, 2021b).



Gambar 7. Alur proses data mining

Dataset yang telah di transformasi selanjutnya dibagi menjadi data training dan data testing (Ferdiana, 2020);(Aldino, Saputra, et al., 2021). Data training bertujuan untuk melatih dataset agar algoritma mengenali mana data yang termasuk ke dalam kelas positif dan mana yang termasuk ke dalam kelas negative (Kurniawan & Susanto, 2019);(Alita et al., 2021). Sesudah dilatih, selanjutnya akan dilakukan tes pada data testing terhadap model yang telah didapat dari data training tersebut. Pada penelitian ini untuk mencari akurasi terbaik maka splitting data dibagi menjadi 4 skenario yaitu membagi dataset secara acak menjadi data training dan data testing dengan rasio sebagai berikut : Pertama : 60% data training dan 40% data testing Kedua : 70% data training dan 30% data testing Ketiga : 80% data training dan 20% data testing Keempat : 90% data training dan 10% data testing.

| Skenario | Training | Testing |
|----------|----------|---------|
| 60 – 40 | 2311 | 1541 |
| 70 – 30 | 2697 | 1155 |
| 80 – 20 | 3081 | 771 |
| 90 – 10 | 3467 | 385 |

Setelah splitting data dilakukan maka selanjutnya adalah melakukan training terhadap algoritma SVM dengan menggunakan data training hasil dari splitting data, setelah itu hasil training tersebut bisa langsung diterapkan untuk klasifikasi sentimen menggunakan data testing (Nurkholis & Sitanggang, 2019);(Technology et al., 2017);(Asia & Samanik, 2018). Untuk mencari akurasi terbaik maka dilakukan perbandingan kernel dari SVM antara lain yaitu kernel linear, rbf, sigmoid, dan polynomial. E. Evaluation Tahapan ini merupakan

tahap akhir dari KDD, untuk memudahkan perhitungan akurasi hasil testing dari setiap dataset pada tahap sebelumnya, maka digunakanlah confusion matrix. Akurasi menunjukkan banyaknya data yang benar diklasifikasikan sesuai label.

HASIL DAN PEMBAHASAN

Dari proses KDD yang telah dilakukan, data yang didapat dari proses scraping pada Twitter mulai dari bulan januari 2019 hingga bulan januari 2020 yaitu sebanyak 3852 yang terbagi menjadi 2034 kelas positif, dan 1818 kelas negative. Setelah dataset memiliki label, langkah selanjutnya data data tersebut akan ditransformasi menjadi data numerik agar bisa diolah oleh algoritma SVM.

| Rasio <i>Splitting</i> <i>Data</i> | Kernel | | | |
|---------------------------------------|--------------|--------|---------|------------|
| | Linear | RBF | Sigmoid | Polynomial |
| 60 : 40 | 0.9637 | 0.9176 | 0.8735 | 0.8884 |
| 70 : 30 | 0.9671 | 0.9333 | 0.8736 | 0.8952 |
| 80 : 20 | 0.9741 | 0.9351 | 0.8677 | 0.8833 |
| 90 : 10 | 0.987 | 0.9325 | 0.839 | 0.8909 |

dapat dilihat bahwa nilai akurasi terendah terdapat pada kernel sigmoid dengan rasio data 90 : 10 yang mendapatkan nilai akurasi sebesar 0.839. Sementara nilai akurasi terbesar terdapat pada kernel linear dengan rasio data 90 : 10 yang mendapatkan nilai akurasi 0.987. Dengan hasil akurasi yang sangat besar yaitu mencapai 98.7% pada penelitian ini membuktikan bahwa algoritma SVM dan teknik skoring untuk labelling memiliki performa yang sangat baik untuk mengklasifikasi data sentimen positif atau negatif. Dari penelitian yang sudah dilakukan dapat dievaluasi bahwa tingkat akurasi ditentukan oleh proses preprocessing, semakin bersih data yang akan diolah semakin bagus juga algoritma SVM dalam mengklasifikasi data. Selain itu hal berpengaruh lainnya untuk menghasilkan akurasi yang bagus adalah jumlah data training dan data testing, dan pemilihan kernel.

SIMPULAN

Setelah dataset diberikan label sentimen positif atau negatif, untuk mengklasifikasikan dataset tersebut menggunakan SVM maka dataset tersebut harus ditransformasikan terlebih dahulu menjadi data numerik karena SVM hanya bisa memproses data numerik saja. Proses transformasi menggunakan teknik TF (term frequency) yaitu menghitung frekuensi

kemunculan kata pada tweet. Setelah selesai di transformasi menjadi data numerik selanjutnya dataset di split menjadi data training dan data testing yaitu dengan pembagian rasio data 60:40, 70:30, 80:20, dan 90:10. Setelah splitting data maka proses klasifikasi dengan algoritma SVM sudah dapat dilakukan dengan menggunakan berbagai kernel yang tersedia pada algoritma SVM. Dari sebanyak 16 kali percobaan klasifikasi, dengan 4 data testing dan 4 kernel yang berbeda-beda, nilai akurasi terbaik didapatkan oleh kernel linear dengan rasio data 90% (3467 data) untuk data training dan 10% (385 data) untuk data testing dengan nilai akurasi yang didapat yaitu sebesar 0.987 atau 98.7%.

REFERENSI

Abdul Maulud, K. N., Fitri, A., Wan Mohtar, W. H. M., Wan Mohd Jaafar, W. S., Zuhairi, N. Z., & Kamarudin, M. K. A. (2021). A study of spatial and water quality index during dry and rainy seasons at Kelantan River Basin, Peninsular Malaysia. *Arabian Journal of Geosciences*, 14(2). <https://doi.org/10.1007/s12517-020-06382-8>

Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021a). Implementation of K-means algorithm for clustering corn planting feasibility area in south lampung regency. *Journal of Physics: Conference Series*, 1751(1), 12038.

Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021b). Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *Journal of Physics: Conference Series*, 1751(1). <https://doi.org/10.1088/1742-6596/1751/1/012038>

Aldino, A. A., Saputra, A., & Nurkholis, A. (2021). *Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur*. 3(3), 325–330. <https://doi.org/10.47065/bits.v3i3.1041>

Alita, D. (2021a). Multiclass Svm Algorithm For Sarcasm Text In Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 118–128.

Alita, D. (2021b). Multiclass SVM Algorithm for Sarcasm Text in Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 118–128. <https://doi.org/10.35957/jatisi.v8i1.646>

Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter. *Jurnal Tekno Kompak*, 14(2), 86–91.

Alita, D., Sari, I., Isnain, A. R., & Styawati, S. (2021). Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 17–23.

Anestiviya, V., Ferico, A., & Pasaribu, O. (2021). Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar). *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(1), 80–85.
<http://jim.teknokrat.ac.id/index.php/JTSI>

Anggraini, S. P., & Suaidah, S. (2022). Sistem Informasi Sentral Pelayanan Publik dan Administrasi Kependudukan Terpadu dalam Peningkatan Kualitas Pelayanan Kepada Masyarakat Berbasis Website (Studi Kasus: Desa Endang Mulyo). *Jurnal Teknologi Dan Sistem Informasi*, 3(1), 12–19.

Anggriani, F., & Husna, N. (2019). Jurnal Manajemen Universitas Bung Hatta Jurnal Manajemen Universitas Bung Hatta. *Jurnal Manajemen Universitas Bung Hatta*, 14(1), 13–19.

Asia, J., & Samanik. (2018). Dissociative Identity Disorder Reflected in Frederick Clegg ' S Character in the Collectors Novel. *ELLiC*, 2(1), 424–431.

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131–145.

Darwis, D., Sulistiani, H., Isnain, A. R., Yasin, I., Hamidy, F., & Mega, E. D. (2022). *Pelatihan pengarsipan secara elektronik (e-filling) bagi perangkat desa di pekon sukanegeri jaya*. 3(1), 108–113.

Dheara, K., Saniati, & Neneng. (2022). *APLIKASI E-COMMERCE UNTUK PEMESANAN SPAREPART MOTOR*. 3(1), 83–89.

Ferdiana, R. (2020). A Systematic Literature Review of Intrusion Detection System for Network Security: Research Trends, Datasets and Methods. *2020 4th International*

Conference on Informatics and Computational Sciences (ICICoS), 1–6.

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>

Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>

Herlinda, V., Darwis, D., & Dartono, D. (2021). Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 94–99.

Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1), 29–36. <https://doi.org/10.31294/jki.v7i1.1>

Isnain, A. R., Hendrastuty, N., Andraini, L., Studi, P., Informasi, S., Indonesia, U. T., Informatika, P. S., Indonesia, U. T., Studi, P., Komputer, T., Indonesia, U. T., & Lampung, K. B. (2021). *Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis*. 6(1), 56–60.

Kurniawan, I., & Susanto, A. (2019). Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019. *Eksplora Informatika*, 9(1), 1–10. <https://doi.org/10.30864/eksplora.v9i1.237>

Lestari, F. P. A., Pane, E. S., Suprpto, Y. K., & Purnomo, M. H. (2018). Wavelet based-analysis of alpha rhythm on eeg signal. *2018 International Conference on Information and Communications Technology (ICOIACT)*, 719–723.

Lina, L. F., & Nani, D. A. (2020). Kekhawatiran Privasi Pada Kesuksesan Adopsi FLina, L. F., & Nani, D. A. (2020). Kekhawatiran Privasi Pada KesukLina, L. F., & Nani, D. A. (2020). Kekhawatiran Privasi Pada Kesuksesan Adopsi FLina, L. F., & Nani, D. A. (2020). Kekhawatiran Privasi Pada Kes. *Performance*, 27(1), 60–69.

Lina, L. F., Nani, D. A., & Novita, D. (2021). *Millennial Motivation in Maximizing P2P*

Lending in SMEs Financing. September, 188–193.

Middleton, E., Kandaswami, C., & Theoharides, T. C. (2000). *The Effects of Plant Flavonoids on Mammalian Cells : Implications for Inflammation , Heart Disease ,.* 52(4), 673–751.

Mittal, A., & Patidar, S. (2019). Sentiment analysis on twitter data: A survey. *ACM International Conference Proceeding Series*, 91–95. <https://doi.org/10.1145/3348445.3348466>

Nurkholis, A., Alita, D., & Munandar, A. (2022). Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(2), 227–233.

Nurkholis, A., & Sitanggang, I. S. (2019). *A spatial analysis of soybean land suitability using spatial decision tree algorithm. December*, 65. <https://doi.org/10.1117/12.2541555>

Prasetyawan, P., Ahmad, I., Borman, R. I., Pahlevi, Y. A., & Kurniawan, D. E. (2018). Classification of the Period Undergraduate Study Using Back-propagation Neural Network. *2018 International Conference on Applied Engineering (ICAE)*, 1–5.

Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jdmsi*, 2(1), 31–37. <https://t.co/NfhnfMjtXw>

Ramadhan, D. A., & Setiawan, B. E. S. S. (2019). Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine. *E-Proceeding of Engineering*, 6(2), 9736–9743.

Septilia, H. A., Parjito, P., & Styawati, S. (2020). Sistem Pendukung Keputusan Pemberian Dana Bantuan menggunakan Metode AHP. *Jurnal Teknologi Dan Sistem Informasi*, 1(2), 34–41.

Styawati, Andi Nurkholis, Zaenal Abidin, & Heni Sulistiani. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 904–910. <https://doi.org/10.29207/resti.v5i5.3380>

Styawati, S., & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 219–230.

Sulistiani, H., & Tjahyanto, A. (2016). Heterogeneous feature selection for classification of customer loyalty fast moving consumer goods (Case study: Instant noodle). *Journal of Theoretical and Applied Information Technology*, 94(1), 77–83.
<https://doi.org/10.5281/zenodo.579836>

Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67.
<https://doi.org/10.47080/simika.v5i1.1411>

Technology, A. I., Indonesia, U. T., & Science, C. (2017). *ALTERNATIVE MODEL BASE AS AN ENABLER FOR SUCCESS OF BUSINESS INTELLIGENCE-BASED*. 95(14).

Tuhuteru, H. (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine. *Information System Development (ISD)*, 5(2), 7–13.

Wahyono, Wibowo, M. E., Ashari, A., & Putra, M. P. K. (2021). Improvement of Deep Learning-based Human Detection using Dynamic Thresholding for Intelligent Surveillance System. *International Journal of Advanced Computer Science and Applications*, 12(10), 472–477. <https://doi.org/10.14569/IJACSA.2021.0121053>

Windane, W. W., & Lathifah, L. (2021). E-Commerce Toko Fisago.Co Berbasis Android. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 2(3), 285–303.
<https://doi.org/10.33365/jatika.v2i3.1139>