



Analisis Sentimen Pengguna Aplikasi E-Wallet Dana Melalui Postingan di Media Sosial Twitter Menggunakan Naïve Bayes

Oktavianus The Memory Hondro

Universitas Islam Sumatera Utara (UISU), Indonesia, email: oktahondro@gmail.com

Info Artikel

Kata Kunci:

Analisis Sentimen; DANA; e-wallet; Twitter; Navie Bayes

Keywords:

Sentiment Analysis; DANA; e-wallet; Twitter; Navie Bayes



Lisensi: cc-by-sa

Copyright © 2023 Oktavianus The Memory Hondro

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 penggunaanya 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.

Abstract

As time goes by, currently information technology is developing increasingly rapidly in every field, one of which is in the financial sector. Now transactions are no longer done manually, starting from ATM machines, mobile banking to the current trend, e-wallet. Transactions using e-wallet are very easy, we only need to exchange the cellphone number that we have registered in the e-wallet application. The DANA application is one of the most popular e-wallet applications because of its good appearance and easy way to use it. However, even though it is popular with users, there are positive and negative opinions. Usually users choose to express opinions openly via social media such as Twitter because they feel more open than expressing opinions directly. To find out whether the DANA application was successful or not, sentiment analysis research was conducted on the Dana application on Twitter social media using the Naïve Bayes method.

1. PENDAHULUAN

Uang elektronik adalah alat pembayaran yang saat ini marak digunakan karena fungsinya yang fleksibel. 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 [1], [2]. Karena maraknya penggunaan uang elektronik untuk mencegah terjadinya kejahatan maka dikembangkan lah aplikasi dompet elektronik atau yang biasa disebut e- wallet [3], [4]. E-wallet ini berfungsi sebagai tempat penyimpanan uang elektronik yang berbentuk aplikasi [5]. Pada e-wallet ini terdapat juga fitur-fitur seperti transfer, pembayaran, pembelian pulsa, token listrik dan lainnya yang sangat membantu memenuhi kebutuhan sehari-hari.

Karena kebutuhan masyarakat yang cukup tinggi dalam penggunaan e-wallet berbagai macam jenis e-wallet diciptakan dengan fitur-fitur unggulan tertentu [6]. 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 [7], [8].

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 [9]. Pada twitter, penggunanya dapat menuangkan pikirannya pada fitur tweet. Oleh karena itu, penelitian analisis sentiment ini menggunakan media sosial twitter karena dinilai lebih efektif dan mudah untuk mengambil data dari setiap tweet pengguna [10].

Dalam melakukan analisis sentimen dibutuhkan metode teks mining, pada penelitian ini, metode yang digunakan adalah metode Naïve Bayes Classifier dengan algoritma (SVM) ([11], [12]. Tujuan penelitian analisis sentimen ini dilakukan untuk mengetahui bagaimana tanggapan pengguna aplikasi DANA terhadap layanan DANA pada sosial media twitter .

2. METODE PENELITIAN

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

1. Selection Data

Pada tahapan awal ini akan dilakukan pencarian dan pengambilan data tweet dari Twitter yang relevan dengan penelitian [13]. 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

2. Preprocessing

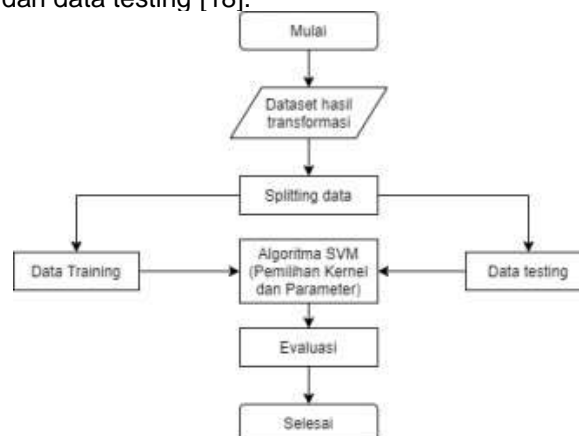
Setelah data tweet tersebut selesai diseleksi, tahapan selanjutnya yaitu tahapan preprocessing [14], [15]. Tahapan preprocessing bertujuan untuk membersihkan data yang kotor. Pada tahapan ini URL, mention, hashtag, angka, simbol, dan tanda baca akan dibersihkan. Selanjutnya mengubah kata-kata yang disingkat 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 .

3. 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 [16]. 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 [17]. 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.

4. Data Mining

Tahapan data mining merupakan tahapan penerapan algoritma untuk klasifikasi. Pada tahapan ini sebelum mulai melakukan penerapan algoritma, kita harus membagi terlebih dahulu dataset menjadi data training dan data testing [18].



Gambar 1. Alur Proses Data Mining

Dataset yang telah di transformasi selanjutnya dibagi menjadi data training dan data testing [19]. 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 ([20]. 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.

Tabel 1. Nilai Skenario, Training dan 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 [21], [22]. 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.

3. HASIL DAN ANALISIS

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.

Tabel 2. Nilai Rasio Kernel

Rasio Splitting Data	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 scoring 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.

4. KESIMPULAN

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

- [1] W. W. Windane and L. Lathifah, "E-COMMERCE TOKO FISAGO.CO BERBASIS ANDROID," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 2, no. 3, pp. 285–303, Oct. 2021, doi: 10.33365/jatika.v2i3.1139.
- [2] D. Darwis, N. Siskawati, and Z. Abidin, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *Jurnal Tekno Kompak*, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.
- [3] L. F. Lina, D. A. Nani, and D. Novita, "MILLENNIAL MOTIVATION IN MAXIMIZING P2P LENDING IN SMEs FINANCING," *JOURNAL OF APPLIED BUSINESS ADMINISTRATION*, vol. 5, no. 2, pp. 188–193, Sep. 2021, doi: 10.30871/jaba.v5i2.3175.
- [4] E. Middleton and C. Kandaswami, "The impact of plant flavonoids on mammalian biology: implications for immunity, inflammation and cancer," in *The Flavonoids*, Springer US, 1993, pp. 619–652. doi: 10.1007/978-1-4899-2911-2_15.
- [5] K. Musanna, R. Sholihin, and M. Sari, "PENGUNAAN E-WALLET OVO PERSPEKTIF ULAMA SYAFI'YAH," *Asy-Syari'ah*, vol. 24, no. 1, pp. 109–120, Jun. 2022, doi: 10.15575/as.v24i1.18073.
- [6] D. Alita, "Multiclass SVM Algorithm for Sarcasm Text in Twitter," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 1, pp. 118–128, Mar. 2021, doi: 10.35957/jatisi.v8i1.646.
- [7] H. A. Septilia, P. Parjito, and S. Styawati, "SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN DANA BANTUAN MENGGUNAKAN METODE AHP," *Jurnal Teknologi dan Sistem Informasi*, vol. 1, no. 2, pp. 34–41, Dec. 2020, doi: 10.33365/jtsi.v1i2.369.
- [8] H. Setiawan, E. Utami, and S. Sudarmawan, "Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 43–51, Jul. 2021, doi: 10.31603/komtika.v5i1.5189.
- [9] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.
- [10] A. Mittal and S. Patidar, "Sentiment Analysis on Twitter Data," in *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, ACM, Jul. 2019. doi: 10.1145/3348445.3348466.
- [11] H. Syah and A. Witanti, "ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 5, no. 1, pp. 59–67, Feb. 2022, doi: 10.47080/simika.v5i1.1411.
- [12] A. Nurkholis, D. Alita, and A. Munandar, "Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 227–233, Apr. 2022, doi: 10.29207/resti.v6i2.3906.
- [13] W. -, Moh. E. Wibowo, A. Ashari, and M. P. K. Putra, "Improvement of Deep Learning-based Human Detection using Dynamic Thresholding for Intelligent Surveillance System," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, 2021, doi: 10.14569/ijacsa.2021.0121053.
- [14] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM," *Jurnal Data Mining dan Sistem Informasi*, vol. 2, no. 1, p. 31, Feb. 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [15] D. Darwis *et al.*, "PELATIHAN PENGARSIPAN SECARA ELEKTRONIK (E-FILLING) BAGI PERANGKAT DESA DI PEKON SUKANEGERI JAYA," *Journal of Social Sciences and Technology for Community Service (JSSTCS)*, vol. 3, no. 1, p. 108, Mar. 2022, doi: 10.33365/jsstcs.v3i1.1946.
- [16] K. N. A. Maulud, A. Fitri, W. H. M. W. Mohtar, W. S. W. M. Jaafar, N. Z. Zuhairi, and M. K. A. Kamarudin, "A study of spatial and water quality index during dry and rainy seasons at Kelantan River Basin, Peninsular Malaysia," *Arabian Journal of Geosciences*, vol. 14, no. 2, Jan. 2021, doi: 10.1007/s12517-020-06382-8.
- [17] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150–155, Oct. 2021, doi: 10.30591/jpit.v6i3.2870.
- [18] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 3, pp. 394–401, Oct. 2018, doi: 10.30591/jpit.v3i3.977.
- [19] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *J Phys Conf Ser*, vol. 1751, no. 1, p. 12038, Jan. 2021, doi: 10.1088/1742-6596/1751/1/012038.
- [20] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naive Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Informatika*, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.

- [21] A. Nurkholis, D. Alita, and A. Munandar, "Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 227–233, Apr. 2022, doi: 10.29207/resti.v6i2.3906.
- [22] T. Wulandari and S. Samanik, "Dissociative Identity Disorder And Its Significance To Nina Sayers Personality Development," *Journal of Arts and Education*, vol. 3, no. 1, Jan. 2023, doi: 10.33365/jae.v3i1.200.