

SENTIMEN ANALISIS MARKETPLACE DI INDONESIA MENGGUNAKAN ALGORITMA RANDOM FOREST

Irgi Mahendrata ¹⁾, Nanda Dwi Husna Sadikin ²⁾, Nandi Dwi Husna Sadikin ³⁾, Niki Marco ⁴⁾,
Yudi Ramdhani ⁵⁾

^{1),2),3),4),5)} Universitas Ardhirajasa Reswara Sanjaya
email : nandwhs03@gmail.com

Abstraksi

Situs marketplace online merupakan tempat belanja yang saat ini sedang populer dikalangan masyarakat, banyak sekali berbagai marketplace yang bersaing, mulai dari shoope, tokopedia, bukalapak dan lain lain. selain memudahkan masyarakat untuk belanja namun ada banyak sekali masalah masalah yang timbul keluhan tentang aplikasi tersebut. Ada berbagai macam masalah pada market place di Indonesia, salah satunya yaitu kurang puasnya pelanggan terhadap market place, yang menyebabkan banyak nya komplain pelanggan terhadap market place tersebut, maka dari itu kami melakukan penelitian analis sentimen melalui data post twitter. Melalui teknologi machine learning, sentimen publik terhadap market place dapat diketahui dengan berbagai macam algoritma, salah satu algoritmanya yaitu algoritma random forest. Penelitian ini bertujuan untuk mengklasifikasi dan mengetahui tingkat akurasi sentimen publik terhadap berbagai macam market place menggunakan teknologi machine learning dengan algoritma random forest. Setelah dilakukan proses analisa dan klasifikasi menggunakan tools rapidminer, tingkat akurasi dari sentimen publik terhadap market place sebesar 68,56%.

Kata Kunci : Marketplace, Random Forest, Sentimen, Twitter.

Abstract

Online marketplace sites are shopping places that are currently popular among the public, there are so many various competing marketplaces, ranging from shoope, tokopedia, bukalapak and others. in addition to making it easier for people to shop but there are a lot of problems that arise complaints about the application. There are various kinds of problems in the market place in Indonesia, one of which is the lack of customer satisfaction with the market place, which causes a lot of customer complaints about the market place, therefore we conducted sentiment analysis research through twitter post data. Through machine learning technology, public sentiment towards the market place can be known with various algorithms, one of which is the random forest algorithm. This research aims to classify and determine the accuracy level of public sentiment towards various market places using machine learning technology with random forest algorithms. After the analysis and classification process using rapidminer tools, the accuracy level of public sentiment towards the market place is 68.56%.

Keywords : Marketplace, Random Forest, Sentiment, Twitter.

PENDAHULUAN

Marketplace sudah menjadi bagian dari kehidupan masyarakat Indonesia saat ini[1]. Marketplace memudahkan masyarakat mencari barang yang diinginkan dan juga memudahkan pedagang untuk mendapatkan pembeli. Pembeli dapat memperoleh barang yang diinginkan tanpa harus keluar rumah. Akan tetapi ada juga masyarakat yang tidak puas dengan marketplace dan terdapat beberapa oknum yang menyalahgunakan marketplace untuk menipu konsumen. Terdapat juga barang-barang yang memakan waktu yang lama untuk sampai ke tangan pembeli dan lewat dari waktu yang dijanjikan.

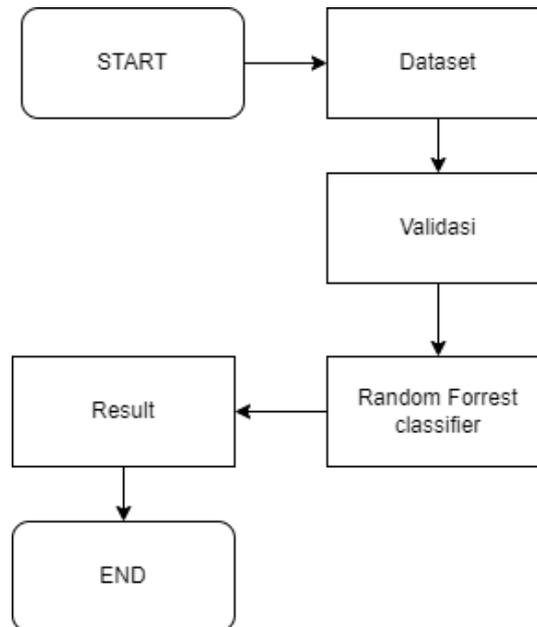
Masyarakat mulai tidak percaya dengan *marketplace* dan penjual, sehingga mereka pun bahkan harus merekam video unboxing sebagai bukti bila barang ditemukan kecacatan ataupun ketika barang tidak sesuai dengan yang dipesan[2]. Selain pembeli, penjual juga mengalami kesulitan karena masyarakat mulai hilang kepercayaan terhadap penjual. Penjual harus pintar mencari jasa pengiriman paket yang cepat dan terpercaya agar tidak mengecewakan konsumen [3]. Sebagian masyarakat memilih tetap melakukan belanja offline untuk menghindari penipuan dan mendapatkan barang dengan lebih cepat [4]. Untuk mengatasi permasalahan ini kami penulis. menggunakan analisis sentimen untuk menemukan pendapat pengguna *twitter* terhadap *marketplace* dan menentukan *marketplace* mana yang dapat dikategorikan baik.

Twitter adalah media sosial yang memungkinkan penggunanya untuk menulis, berbagi dan membaca pesan teks yang disebut tweet [5]. *Twitter* dapat menjadi sumber daya yang potensial untuk mengumpulkan data yang dapat dianalisa dan diklasifikasi. Data pengguna dari *Twitter* akan diperoleh dengan menggunakan teknik *web scraping* yang merupakan proses pengambilan data dari halaman web yang dituju untuk mengambil informasi yang dibutuhkan dengan menggunakan perangkat yang disebut *crawler* [6] .

Analisis sentimen adalah proses untuk mengidentifikasi opini dalam teks, yang dapat berupa opini positif, negatif, atau netral. Dalam penelitian ini, kami menggunakan metode *Random Forest*, yang merupakan metode input untuk *Decision tree*. *Decision tree* digunakan untuk mengklasifikasikan sampel data yang tidak dikenal ke dalam kelas yang ada. Pada penelitian sebelumnya, analisis sentimen telah dilakukan oleh Bern Jonathan, Jay Idoan Sihotang, dan Stanley Martin [7], yang mengkategorikan kelas sentimen sebagai positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa penelitian ini memiliki tingkat presisi sebesar 92%, presisi positif sebesar 92%, presisi negatif sebesar 93%, presisi netral sebesar 96%, recall positif sebesar 99%, recall negatif sebesar 89%, dan recall netral sebesar 73%. Rata-rata presisi adalah 93% dan recall adalah 87%. Berdasarkan penjelasan ini, kami memiliki ide untuk melakukan analisis sentimen terhadap opini pengguna terhadap *marketplace* di Indonesia. Dataset diperoleh dari tweet pengguna di media sosial *Twitter*. Penulis menggunakan algoritma *random forest* untuk mengklasifikasikan opini pengguna *marketplace* Indonesia dari tweet di media sosial *Twitter* menjadi tiga kelas sentimen, yaitu positif, netral, dan negatif. Metrik yang digunakan untuk mengevaluasi hasil klasifikasi adalah Presisi, Akurasi, recall, dan skor F1. Algoritma *random forest* dipilih karena dianggap cocok untuk klasifikasi data dengan sampel yang besar. Penelitian ini diharapkan dapat digunakan oleh perusahaan *marketplace* untuk meningkatkan kualitas layanan dan mempertahankan fitur yang dinilai positif oleh pengguna.

METODE PENELITIAN

Dalam penelitian ini, kami mengevaluasi keefektifan algoritma *Random Forest* dalam pengklasifikasian data. Tahap-tahap yang dilakukan dalam penelitian ini meliputi pengumpulan dataset, validasi, pemilihan fitur, pembuatan model, evaluasi model, dan perbaikan model jika diperlukan. Pertama, data yang digunakan dalam penelitian ini diperoleh dari sumber tertentu. Kemudian, data yang diperoleh divalidasi untuk mengevaluasi kualitas dan keakuratan data. Selanjutnya, fitur-fitur yang digunakan dalam pengklasifikasian ditentukan melalui proses seleksi fitur. Setelah fitur-fitur yang digunakan ditentukan, model *Random Forest* dibangun dengan menggunakan data yang telah disiapkan. Model yang dibangun kemudian diuji dengan menggunakan data uji untuk mengevaluasi performa model. Metrik seperti akurasi, presisi, dan recall digunakan untuk mengevaluasi model. Jika diperlukan, model dapat diperbaiki dengan mengubah parameter atau dengan menambah data latih. Berdasarkan hasil evaluasi model, ditarik kesimpulan tentang keefektifan algoritma *Random Forest* dalam pengklasifikasian data yang digunakan dalam penelitian ini. Terdapat beberapa tahapan pada gambar 1, mencakup: Pengumpulan dataset, validasi, pengklasifikasian data menggunakan algoritma *random forest* dan hasil.



Gambar 1. Alur Penelitian

A. PENGUMPULAN DATA

Data diambil dari media sosial twitter sebanyak 1165 post mengenai berbagai marketplace di Indonesia. Data berisi ulasan, keluhan, diskusi, dan evaluasi tentang marketplace dari sisi konsumen maupun penjual. Data dikumpulkan dari post sekitar November 2022 hingga Januari 2023. Data dikumpulkan oleh 12 orang dengan metode pengumpulan data yang berbeda beda dan kata kunci yang berbeda. Dari data twitter yang telah dikumpulkan, kemudian dikelompokkan menjadi tiga kelas yaitu postingan positif, negative dan netral. Hasil dari pengelompokan tiga kelas tersebut adalah 798 data postingan positif, 268 postingan negative dan 99 postingan netral. Dataset ini digunakan untuk melakukan analisis dan klasifikasi sentiment *twitter* tentang *marketplace* yang ada di Indonesia.

Tabel 1. Distribusi Sentimen Tweet Marketplace

Status Sentimen	Count
Positive	798
Negative	268
Netral	99

B. VALIDASI

Kumpulan sampel secara acak dibagi menjadi serangkaian lipatan (kelompok) dengan ukuran yang sama menggunakan teknik *cross validation k-fold*, di mana k adalah jumlah lipatan (partisi) dalam kumpulan data[8]. Kami menggunakan *cross validation 10-fold* yang artinya kami membagi dataset menjadi sepuluh partisi. Sembilan partisi digunakan untuk data pelatihan, sedangkan partisi yang tersisa digunakan untuk data pengujian. Menggunakan partisi yang berbeda sebagai set pengujian dan sembilan partisi yang tersisa sebagai data pelatihan, pelatihan diulang sepuluh kali. Setelah itu, rata-rata hasilnya disajikan.

C. KLASIFIKASI MEMAKAI ALGORITMA RANDOM FOREST

Klasifikasi adalah proses menemukan model dari data. Tujuan dari klasifikasi adalah untuk membuat keputusan dengan memprediksi suatu kasus berdasarkan hasil klasifikasi yang didapatkan[9]. *Random forest* adalah sekelompok pengklasifikasi yang terdiri dari banyak pohon keputusan dan mengklasifikasikan berdasarkan output dari hasil klasifikasi dari setiap anggota pohon keputusan[10]. Dengan penggunaan pohon-pohon tersebut akan meningkatkan akurasi yang akan diperoleh akan menjadi lebih baik.

Ketika proses klasifikasi sudah selesai, inialisasi dilakukan dengan data sebanyak-banyaknya berdasarkan nilai akurasi. Keuntungan dari penggunaan *random forest* adalah dapat mengklasifikasikan data yang memiliki atribut yang tidak lengkap, dapat digunakan untuk klasifikasi dan *regresi* tetapi tidak terlalu baik untuk *regresi*, lebih cocok untuk mengklasifikasikan data dan dapat digunakan untuk menangani data sampel yang besar. Klasifikasi Proses klasifikasi pada *random forest* dimulai dari memecah data sampel yang ada ke dalam pohon-pohon keputusan acak. Setelah pohon tersebut terbentuk, maka akan dilakukan voting pada setiap kelas dari sampel data sampel. Kemudian, menggabungkan suara dari setiap kelas dan kemudian mengambil suara terbanyak.

Menggunakan metode *random forest* pada proses klasifikasi data akan menghasilkan hasil yang optimal. Teknik ini dibangun dengan menggunakan bagging dan pemilihan atribut yang dilakukan secara acak. Metode *CART (Classification and Regression Tree)* digunakan untuk menumbuhkan pohon keputusan yang tidak dipangkas sehingga menghasilkan kumpulan pohon yang disebut forest. Adapun kelebihan dari metode *Forest* sebagai berikut [11]:

- Hasil akurasi lebih baik

Random forest menunjukkan kinerja yang lebih baik dalam hal akurasi dibandingkan dengan metode klasifikasi lainnya seperti *single decision tree*, *k-NN*, atau *SVM*. Ini karena teknik *ensemble learning* yang digunakan, yaitu menggabungkan prediksi dari beberapa pohon keputusan yang berbeda dan kemudian melakukan *voting* untuk menghasilkan prediksi akhir.

- Lebih kuat dalam outliers dan noise

Random forest mampu menangani data yang mengandung *outliers* atau *noise* karena penggunaan teknik *bootstrapping* dan pembangunan pohon keputusan dari sampel acak dari data asli.

- Lebih cepat daripada bagging dan boosting

Random forest memiliki kecepatan yang lebih cepat dibandingkan dengan metode *bagging* atau *boosting* karena setiap pohon keputusan dapat dibangun secara paralel. Selain itu, *random forest* menggunakan teknik pemilihan atribut secara acak yang mempercepat proses pelatihan.

- Mempunyai sifat yang sederhana dan mudah dipararelkan

Random forest memiliki sifat yang sederhana dan mudah diimplementasikan, karena hanya membutuhkan beberapa parameter yang harus diatur dan dapat dilakukan secara paralel. Selain itu, *random forest* mudah diinterpretasikan karena hasilnya dapat divisualisasikan dalam bentuk pohon keputusan.

D. EVALUASI

Confusion Matrix adalah metode yang digunakan untuk mengevaluasi kinerja model klasifikasi. Matriks ini menunjukkan jumlah prediksi benar dan salah yang dibuat oleh model untuk setiap kelas. Dari *matriks* ini, kita dapat menghitung beberapa metrik penting seperti akurasi, presisi, dan daya ingat [12]. Akurasi adalah rasio jumlah prediksi yang benar dengan jumlah total prediksi. Presisi adalah rasio jumlah prediksi positif yang sebenarnya terhadap jumlah total prediksi positif. *Recall* adalah rasio jumlah prediksi positif yang benar dengan jumlah kelas positif yang sebenarnya. Dengan menggunakan *Matriks Kebingungan*, kami dapat mengevaluasi kinerja model klasifikasi secara lebih rinci dan mengidentifikasi area yang memerlukan perbaikan. *Confusion Matrix* akan digunakan untuk menentukan akurasi, *precision*, dan *recall*. Dengan rumus perhitungan sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Tabel 2. Confusion matrix

Pada gambar 5 dapat dilihat bahwa pengujian menggunakan algoritma *Random Forest* menunjukkan hasil yang tidak sangat baik dalam hal akurasi. Nilai akurasi yang diperoleh dari pengujian ini adalah 68,56%. Selain itu, nilai precision dari prediksi positif, negatif, dan netral adalah 68,56%, 0,00%, dan 0,00%. Recall dari masing-masing prediksi juga rendah, yaitu 100,00%, 0,00%, dan 0,00%. Ini menunjukkan bahwa algoritma *Random Forest* tidak memiliki tingkat akurasi yang tinggi. Pengujian ini dilakukan dengan *metode Cross Validation* dengan *10-Fold*.

Dilihat dari hasil pengujian tersebut diketahui bahwa algoritma *Random Forest* memiliki tingkat akurasi yang tidak terlalu tinggi. Hasil dari pengujian menunjukkan jumlah akurasi dari algoritma dengan nilai *precision* dan *recall*. Adapun dalam pengujian algoritma yang telah dilakukan diatas menggunakan *metode Cross Validation* dengan *10-Fold*.

KESIMPULAN

Penelitian ini dilakukan untuk mengklasifikasikan analisis sentimen pada *MarketPlace* di Indonesia menggunakan algoritma random forest. Data sentimen alisis yang telah di kumpulkan dari media sosial *twitter* yang telah dikumpulkan, kemudian dikelompokan menjadi tiga kelas yaitu postingan positif, negative dan netral. Hasil dari pengelompokan tiga kelas tersebut adalah 798 data postingan positif, 268 postingan negative dan 99 postingan netral. Hasil dari analisis sentiment masyarakat terhadap *MarketPlace* di Indonesia memiliki beragam sentimen, baik itu sentiment positif maupun sentimen negatif. Hasil klasifikasi dengan menggunakan algoritma *Random Forest* untuk analisis sentimen publik terhadap marketplace yang kami lakukan menghasilkan nilai akurasi sebesar 68,56%. Hasil tersebut membuktikan bahwa model *Random Forest* memiliki kinerja yang tidak cukup baik dalam melakukan analisis sentimen. Dari hasil yang didapatkan dapat disimpulkan bahwa algoritma *Random forest* memiliki kinerja yang tidak cukup baik untuk mengklasifikasikan sentimen analis *marketplace* di Indonesia. Saran penelitian untuk kedepannya, kami akan mencoba menggunakan algoritma klasifikasi seperti *Algoritma SVM* atau yang lain untuk mendapatkan hasil nilai akurasi yang lebih baik

DAFTAR PUSTAKA

- [1] R. ROHAINI, S. H. Dewi Septiana, and P. Joshua, "Penyelesaian Sengketa Online Marketplace Antara Penjual dan Pembeli Melalui Online Dispute Resolution," *Pactum Law Journal*, vol. 2, no. 01, pp. 537–549, 2018.
- [2] N. N. U. R. CHANIFAH, "Pandangan ekonomi Islam terhadap jual beli melalui marketplace online Shopee".
- [3] A. Prasetio *et al.*, *Konsep Dasar E-Commerce*. Yayasan Kita Menulis, 2021.
- [4] I. U. Yovirizka and W. Wahjoedi, "TREND KEPUTUSAN BELANJA ONLINE MAHASISWA FAKULTAS EKONOMI FE UM ANGKATAN 2014," *Jurnal Pendidikan Ekonomi*, vol. 11, no. 1, pp. 41–47, 2018.
- [5] T. F. Basar, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, vol. 2548, p. 964X.
- [6] S. Kumar, G. Barbier, M. Abbasi, and H. Liu, "Tweetracker: An analysis tool for humanitarian and disaster relief," in *Proceedings of the international aaai conference on web and social media*, 2011, vol. 5, no. 1, pp. 661–662.
- [7] B. Jonathan, J. I. Sihotang, and S. Martin, "Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier," in *Abstract Proceedings International Scholars Conference*, 2019, vol. 7, no. 1, pp. 1719–1728.
- [8] C. A. Ramezan, T. A. Warner, and A. E. Maxwell, "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification," *Remote Sens (Basel)*, vol. 11, no. 2, p. 185, 2019.
- [9] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable sentiment classification for big data analysis using naive bayes classifier," in *2013 IEEE international conference on big data*, 2013, pp. 99–104.
- [10] Y. Mao, Y. He, L. Liu, and X. Chen, "Disease classification based on eye movement features with decision tree and random forest," *Front Neurosci*, vol. 14, p. 798, 2020.
- [11] L. Ratnawati and D. R. Sulistyanningrum, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, pp. A71–A77, 2020.
- [12] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020.