

Detecting Bot Comments on a Product in Shopee Using the Gradient Boosting Method

Manda Sari

Informatics Engineering,
Malikussaleh University,
Aceh, Indonesia
manda.210170103@mhs.unimal.ac.id

Rizal *

Informatics Engineering,
Malikussaleh University,
Aceh, Indonesia
rizal@unimal.ac.id

*Corresponding Author

Sujacka Retno

Informatics Engineering,
Malikussaleh University,
Aceh, Indonesia
sujacka@unimal.ac.id

• Submitted: 2025-05-17; Accepted: 2025-05-31; Published: 2025-06-05

Abstract—Shopee is one of the largest e-commerce platforms in Southeast Asia, providing a product comment feature that serves as a primary reference for prospective buyers to assess product quality and seller reputation. Unfortunately, the prevalence of fake comments generated by bots—characterized by rigid language, repetitive patterns, and excessive praise—raises concerns about the authenticity of available reviews. This issue can negatively influence consumers' purchasing decisions. This study aims to develop an automated system capable of detecting bot comments using the Gradient Boosting algorithm. A total of 3,000 comments were manually collected from various product categories and labeled directly by the researchers. The comment data were then processed through several stages, including text cleaning, tokenization, and lemmatization, to prepare for model analysis. The trained model demonstrated excellent performance, achieving an accuracy of 94.09%, precision of 95.99%, recall of 83.23%, and an F1-score of 89.13%. Based on these results, it can be concluded that the Gradient Boosting algorithm is highly effective in classifying bot comments and can help improve consumer trust and security in online shopping.

Keywords—Bot Comment Detection, Gradient Boosting, E-commerce, Shopee, Text Classification

I. PENDAHULUAN

Perkembangan pesat *e-commerce* di Indonesia, terutama pada platform seperti Shopee, telah menciptakan dinamika baru dalam interaksi antara penjual dan pembeli (Gudiato, Sedyono, and Sembiring 2022). Salah satu elemen penting yang mempengaruhi keputusan pembelian adalah komentar produk (Puspitasari et al. 2024), yang berfungsi sebagai indikator kepercayaan konsumen terhadap kualitas produk atau layanan yang diberikan. Komentar-komentar ini, baik yang positif maupun negatif, membantu calon pembeli untuk menilai apakah produk tersebut sesuai dengan kebutuhan mereka (E H Muktafin, Kusri, and Luthfi 2020).

Namun, kemajuan teknologi juga memunculkan tantangan baru (Rusmarasy, Priyambadha, and Pradana

2019), yaitu penggunaan bot untuk memanipulasi sistem (Lestari 2022) komentar dengan cara yang tidak jujur, seperti memberikan komentar otomatis yang menguntungkan penjual atau menyebarkan pujian palsu (Dwiyanaputra et al. n.d.). Bot sering digunakan untuk meningkatkan rating produk atau memberikan kesan positif secara otomatis, tanpa adanya pengalaman nyata dari pengguna (Priyatno et al. 2019). Komentar yang dihasilkan oleh bot sering kali memiliki ciri-ciri yang dapat diidentifikasi, yang berbeda dengan komentar yang ditulis oleh manusia (Munawar et al. 2023). Salah satu aspek penting dalam mendeteksi komentar *bot* adalah pengenalan pola bahasa, termasuk cara komentar tersebut memuji produk. Pujian yang diberikan oleh bot cenderung mengikuti pola-pola tertentu yang bisa dibedakan dengan menganalisis pola bahasa yang kaku, baku, berulang, dan pujian untuk membedakan komentar bot dan manusia.

Penelitian oleh Silvia Elsa Suryana dan tim menggunakan metode Gradient Boosting untuk memprediksi keberhasilan telemarketing produk perbankan. Dengan 20 faktor sebagai input dan hasil "ya" atau "tidak" sebagai output, model ini dioptimalkan dengan teknik khusus agar menghasilkan performa terbaik. Hasilnya, model ini mencapai akurasi 90,39%, presisi 94,91%, dan skor AUC 0,939, yang menunjukkan kemampuannya memprediksi calon pelanggan dengan sangat baik (Elsa Suryana and Warsito n.d.). Penelitian lain oleh Ridwansyah dan rekan pada tahun 2023 menunjukkan bahwa Gradient Boosting sangat efektif dalam memprediksi risiko diabetes tipe 2. Tingkat akurasi metode ini berkisar antara 81% hingga lebih dari 90%, lebih baik dibandingkan metode lain seperti Random Forest, Naïve Bayes, dan Bagging. Selain akurasi tinggi, Gradient Boosting juga unggul dalam memahami hubungan kompleks antar faktor risiko, sehingga memberikan prediksi yang lebih akurat dan mendalam (Ridwansyah and Zakaria 2023).

Penelitian ini bertujuan mengembangkan sistem deteksi komentar bot di Shopee menggunakan *Gradient Boosting*, dengan pendekatan kuantitatif melalui proses pengumpulan, pelabelan, dan analisis data komentar (Urllamma, Supriya, Lavanya, and Hari Priya 2024).

Implementasi ini diharapkan dapat meningkatkan integritas dan kepercayaan dalam sistem *e-commerce*, memperbaiki kualitas interaksi antara penjual dan pembeli (Muhamad Amhar Rayadin et al. 2024), serta mengurangi dampak negatif dari penggunaan *bot* dalam komentar dan meningkatkan kepercayaan pengguna serta mendukung ekosistem *e-commerce* yang lebih transparan dan terpercaya (Elik Hari Muktafin, Kusri, and Luthfi 2020).

II. KAJIAN LITERATUR

A. Gradient Bosting

Gradient Boosting merupakan metode dalam pembelajaran mesin yang digunakan untuk meningkatkan akurasi model prediktif dengan membangun model secara bertahap (Ullah, Supriya, Lavanya, and Priya 2024). Metode ini bekerja dengan menggabungkan sejumlah model lemah, seperti pohon keputusan berukuran kecil, di mana setiap model baru dirancang untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya. Prosesnya dimulai dengan membuat prediksi awal, menghitung kesalahan prediksi, kemudian membangun model selanjutnya berdasarkan nilai residual (selisih antara prediksi dan label sebenarnya) yang dihasilkan. Gradient atau turunan dari fungsi error digunakan untuk menentukan arah dan besarnya perbaikan pada setiap iterasi (Muhamad Amhar Rayadin et al. 2024). Dengan pendekatan bertahap ini, Gradient Boosting mampu menangani data yang kompleks (Nurdin, Rizal, and Rizwan 2019) dan banyak digunakan dalam berbagai aplikasi seperti deteksi penipuan, analisis teks, dan pengenalan gambar. Dalam penelitian ini, algoritma Gradient Boosting digunakan untuk meningkatkan akurasi klasifikasi (Rosnita n.d.) komentar bot dengan menghasilkan model yang lebih kuat dan andal dari hasil penggabungan beberapa model sederhana.

Rumus Gradient Boosting:

1. Rumus Awal (Inisialisasi Model Awal):

Model awal F^0 adalah rata-rata dari seluruh nilai target y_i . Ini merupakan prediksi awal yang digunakan sebelum proses boosting dimulai. Nilai F^0 dihitung menggunakan formula (1).

$$F_0 = \frac{1}{n} \sum_{i=1}^n y_i \quad (1)$$

2. Menghitung Residual atau Error pada Iterasi ke-t

Residual $r_i^{(t)}$ menunjukkan selisih antara nilai aktual dengan hasil prediksi model sebelumnya. Nilai residual ini yang akan dipelajari oleh model berikutnya yang dihitung menggunakan formula (2).

$$r_i^{(t)} = y_i - F_{t-1}(x_i) \quad (2)$$

3. Menghitung Koefisien Skala (γ)

Koefisien γ digunakan untuk mengalikan hasil prediksi dari model ke-t agar bisa memperbaiki kesalahan dari model sebelumnya secara optimal. Nilai tersebut dihitung menggunakan formula (3).

$$\gamma_t = \frac{\sum_{i=1}^n r_i^{(t)} \cdot h_t(x_i)}{\sum_{i=1}^n h_t(x_i)^2} \quad (3)$$

4. Update Model:

Model diperbarui dengan menambahkan hasil prediksi dari model baru (*weak learner*) yang telah dikalikan dengan koefisien γ_t . Proses ini diulang sampai iterasi selesai menggunakan formula (4).

$$F_t(x) = F_{t-1}(x) + \gamma_t \cdot h_t(x) \quad (4)$$

Keterangan:

F^0 : Prediksi awal (bias) dari model.

n : Jumlah total data.

y_i : Nilai aktual/target dari data ke-i.

$r_i^{(t)}$: Residual (kesalahan) pada data ke-i di iterasi ke-t.

$F_{t-1}(x_1)$: Hasil prediksi model pada iterasi sebelumnya (t-1) terhadap data ke-i.

γ_t : Koefisien skala pada iterasi ke-t.

$h_t(x_1)$: Output model weak learner (misalnya pohon keputusan) ke-t untuk data ke-i.

$F_t(x)$: Model akhir pada iterasi ke-t.

$F_{t-1}(x)$: Model dari iterasi sebelumnya (t-1)

$h_t(x)$: Model baru (weak learner) ke-t.

B. XGBoost (Extreme Gradient Boosting)

XGBoost (*Extreme Gradient Boosting*) adalah algoritma pembelajaran mesin yang sangat populer karena kemampuannya yang cepat dan akurat dalam menyelesaikan masalah klasifikasi dan regresi (Wijaya, Yulianto, and Prasetyo 2021). Ini merupakan pengembangan dari metode Gradient Boosting, yang dirancang untuk memberikan hasil yang lebih baik, terutama dalam kompetisi di bidang data science. Salah satu keunggulan utama XGBoost adalah kecepatan dan efisiensinya. Algoritma ini dapat memanfaatkan sumber daya dengan baik, termasuk menggunakan memori secara efisien dan memproses data secara paralel, sehingga waktu pelatihan menjadi lebih singkat (M Amhar Rayadin, Musaruddin, and Saputra 2024). Selain itu, XGBoost dilengkapi dengan fitur regulasi yang membantu mencegah model dari overfitting (Kurniawan and Raja Ali Haji Jl Politeknik Senggarang 2020), yaitu ketika model terlalu cocok dengan data pelatihan sehingga tidak dapat bekerja dengan baik pada data baru. XGBoost juga mampu menangani data yang hilang secara langsung (Rosnita, Retno, and Hariono n.d.), tanpa perlu mengisi atau menghapusnya terlebih dahulu.

C. Text Preprocessing

Pada Figure 1 adalah langkah penting dalam pengolahan data teks yang bertujuan untuk mempersiapkan data agar dapat digunakan dalam analisis lebih lanjut, seperti klasifikasi, pemodelan, atau analisis sentimen. Proses ini melibatkan serangkaian teknik untuk membersihkan dan menyusun data teks agar lebih terstruktur dan mudah dianalisis.

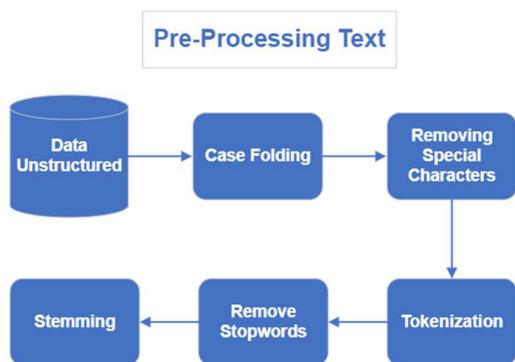


Figure. 1. Pre-processing Text Data

Berikut adalah beberapa langkah umum dalam text preprocessing:

1. *Data Unstructured*:

Respons penjual dari Shopee dikumpulkan. Respons ini dapat berupa berbagai format, mungkin mengandung bahasa campuran, kata-kata informal, atau bahasa gaul, sehingga tidak terstruktur.

2. *Case Folding*:

Ubah semua teks menjadi huruf kecil untuk memperlakukan kata seperti "BELI" dan "beli" secara setara, mengurangi variasi karena penggunaan huruf besar.

3. *Removing Special Characters*:

Banyak respons yang mungkin menyertakan emoji, tanda baca, atau simbol. Menghapus karakter-karakter ini membantu membersihkan data, dengan fokus pada kata-kata yang relevan untuk mendeteksi bot atau respons manusia.

4. *Tokenization*:

Memecah respons menjadi token individual (kata-kata) membantu dalam menganalisis pola, seperti frekuensi kata atau frasa spesifik yang digunakan oleh bot atau manusia.

5. *Remove Stopwords*:

Kata-kata umum yang tidak memberikan makna, seperti "the," "is," atau "dan" (dalam Bahasa Indonesia), dihapus. Langkah ini memungkinkan pemfokusan pada kata kunci yang dapat memberi sinyal apakah respons dihasilkan oleh bot atau manusia.

6. *Stemming*:

Kata-kata disederhanakan menjadi bentuk akarnya. Ini membantu mengidentifikasi kata-kata yang mirip (seperti "terima," "menerima," dan "diterima") sebagai konsep yang sama, yang dapat membantu mendeteksi pola dalam penggunaan bahasa bot vs. manusia.

Dengan melakukan text preprocessing, data teks menjadi lebih bersih dan terstruktur, sehingga memudahkan penerapan teknik analisis lanjutan. Langkah-langkah ini sangat penting dalam memastikan bahwa model yang dibangun dapat memberikan hasil yang akurat dan relevan.

III. METODE

Metode penelitian ini dirancang untuk mengidentifikasi apakah ulasan komentar di Shopee berasal dari manusia atau bot. Proses ini meliputi tiga langkah utama: pengumpulan data, pelabelan, dan pengolahan data.

Masing-masing langkah memiliki peran penting dalam menghasilkan model klasifikasi yang akurat.

A. Pengumpulan Data

Pengumpulan data adalah langkah awal yang sangat penting dalam penelitian ini, karena kualitas data yang dikumpulkan akan sangat mempengaruhi hasil analisis. Penelitian ini fokus pada ulasan dan komentar pembeli terhadap produk di Shopee untuk membedakan komentar asli dari manusia dengan komentar yang dibuat oleh bot.

Proses pengumpulan data dilakukan secara manual dengan mengamati langsung halaman produk di Shopee. Peneliti mencatat komentar satu per satu, mencakup elemen-elemen seperti isi teks komentar, waktu pengiriman komentar, serta informasi tambahan seperti nama produk dan kategori. Pendekatan manual ini memungkinkan peneliti untuk memilih data yang relevan dan mempertimbangkan konteks komentar yang disukai pembeli. Selama proses ini, peneliti juga mencatat faktor-faktor yang mempengaruhi karakteristik komentar, seperti gaya bahasa yang digunakan, keberadaan pujian, dan pola komentar yang cenderung diulang. Data yang terkumpul akan disimpan dalam format terstruktur, seperti file CSV atau database, untuk memudahkan proses analisis lebih lanjut.

Dengan metode pengumpulan data manual, penelitian ini bertujuan untuk menghasilkan dataset berkualitas tinggi yang mendukung pengembangan model klasifikasi menggunakan metode *Gradient Boosting*. Model ini diharapkan mampu mengidentifikasi pola-pola komentar bot, seperti gaya bahasa yang kaku, baku, berisi pujian berulang, dan kurang memiliki variasi, sehingga dapat meningkatkan integritas ulasan produk serta kepercayaan konsumen terhadap interaksi di Shopee.

B. Labelisasi

Setelah data berhasil dikumpulkan, langkah berikutnya adalah proses labelisasi. Labelisasi adalah tahap penting di mana setiap komentar pembeli diberi label untuk menunjukkan apakah komentar tersebut berasal dari manusia atau bot. Proses ini diperlukan untuk menyediakan dataset terlabel yang akan digunakan dalam pengembangan model klasifikasi.

Pada Figure 2 merupakan cara pelabelan data secara otomatis berdasarkan pola teks dalam ulasan (*review*). Setiap fungsi mendeteksi pola tertentu seperti bahasa kaku, komentar berulang, bahasa baku, dan pujian metafora. Hasil deteksi dari masing-masing fungsi digunakan untuk membuat kolom label baru di dataset, seperti kaku, berulang, baku, dan pujian_metafora, yang bernilai 1 jika pola terdeteksi, dan 0 jika tidak. Dengan demikian, proses ini membantu memberi label pada data secara sistematis untuk keperluan analisis atau pelatihan model. Tahapan ini bertujuan untuk menciptakan dataset yang kuat sehingga model mampu mendeteksi pola dan karakteristik unik dari komentar bot meningkatkan akurasi analisis, serta mendukung tujuan penelitian dalam mengidentifikasi komentar yang autentik dan membangun kepercayaan konsumen di Shopee.

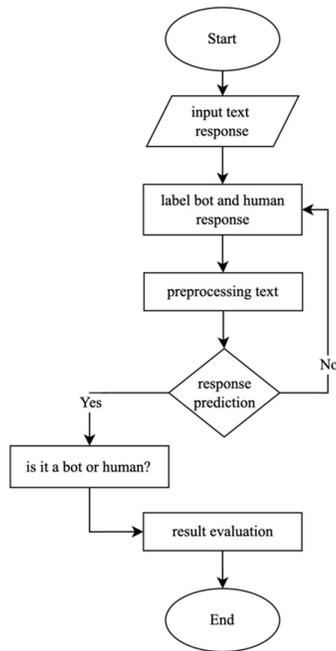


Figure. 3. Flowchart

IV. HASIL DAN PEMBAHASAN

A. Modeling

Pada tahap *modeling*, dilakukan proses pelatihan model machine learning dengan algoritma *Gradient Boosting Classifier* untuk mendeteksi komentar bot pada penjualan produk di Shopee. Proses ini diawali dengan menyiapkan fitur-fitur yang telah diekstraksi dari data komentar, yaitu pola kaku, pola berulang, kebakuan bahasa, dan pujian dengan metafora. Keempat fitur tersebut kemudian digunakan sebagai variabel independen (X), sedangkan label hasil klasifikasi bot atau bukan bot berdasarkan jumlah pola yang terdeteksi digunakan sebagai variabel dependen (y). Dataset yang digunakan sebanyak 3000 data. Pada Table 1 merupakan contoh dataset yang digunakan. Selanjutnya, data dibagi menjadi dua bagian menggunakan teknik *train-test split*, yaitu 80% data untuk pelatihan dan 20% untuk pengujian. Model *Gradient Boosting* kemudian dilatih menggunakan data pelatihan dengan memanggil fungsi *fit()* untuk membangun model prediksi. Setelah proses pelatihan selesai, model yang telah terbentuk disimpan dalam bentuk file menggunakan *joblib.dump()*, sehingga dapat digunakan kembali saat proses prediksi tanpa perlu melakukan pelatihan ulang. Tahapan ini merupakan inti dari proses pengembangan sistem deteksi bot, karena model inilah yang menjadi dasar dalam memprediksi apakah sebuah komentar termasuk komentar bot atau bukan pada Figure 4.

```

45 # .....
46 # Training Model
47 # .....
48 X = df[['kaku', 'berulang', 'baku', 'pujian_metafora']]
49 y = df['label_bot']
50
51 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
52
53 model = GradientBoostingClassifier()
54 model.fit(X_train, y_train)
55

```

Figure. 4. Model Detection Process

Table 1. Shopee Comments Dataset

No	Review	Kaku	Berulang	Baku	Pujian Metafora	Label
1	Alhamdulillah ini sangat bagus,amanah, pas dan cocok, saya akan membeli lagi di took ini	0	1	1	0	1
2	pengiriman lama	0	1	1	0	1
3	bagus sesuai gambar	0	0	1	0	0
4	pengiriman nya lama banget untung bajunya lumayan	0	0	1	0	0
5	Terimakasih telah mengirimkan baju yang sangat sesuai dan akan saya beli lagi di toko ini	0	1	1	0	1
6	ada harga ada kualitas intinya, bajunya sesuai dengan harganya	0	0	1	0	0

B. Evaluasi

Evaluasi dilakukan untuk mengukur seberapa baik model *Gradient Boosting* dalam membedakan komentar bot dan komentar manusia pada platform Shopee. Evaluasi ini dilakukan dengan cara membandingkan hasil prediksi model terhadap data uji yang sudah diketahui kebenarannya.

Dataset dibagi menjadi dua bagian:

1. 80% digunakan sebagai data latih untuk melatih model
2. 20% digunakan sebagai data uji untuk menguji model

Setelah model selesai dilatih, prediksi dilakukan terhadap data uji. Hasil prediksi ini dibandingkan dengan data sebenarnya untuk menghitung performa model menggunakan beberapa ukuran evaluasi, yaitu: akurasi, precision, recall, dan F1-score.

Confusion Matrix digunakan untuk menunjukkan seberapa banyak komentar yang diprediksi benar atau salah oleh model. Hasil *Confusion Matrix* dari model *Gradient Boosting* ditampilkan pada Figure 5 dan 6.

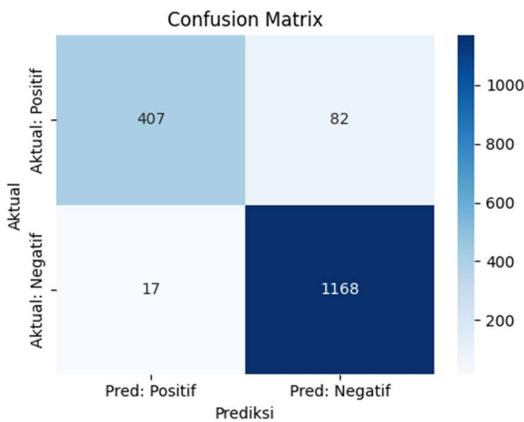


Figure 5. Confusion Matrix

Penjelasan:

True Positive (TP): 407

False Positive (FP): 82

False Negative (FN): 17

True Negative (TN): 1168

Berdasarkan data pada Confusion Matrix, berikut hasil perhitungan dari setiap metrik evaluasi:

a. Akurasi

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{407 + 1168}{407 + 1168 + 82 + 17}$$

$$= \frac{1575}{1674} = 83,15\%$$

b. Precision

$$Precision = \frac{TP}{TP + FP}$$

$$= \frac{407}{407 + 82} = 83,23\%$$

c. Recall

$$Recall = \frac{TP}{TP + FN}$$

$$= \frac{407}{407 + 17} = 95,99\%$$

d. F1-score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$= 2 \times \frac{0,8323 \times 0,9599}{0,8323 + 0,9599}$$

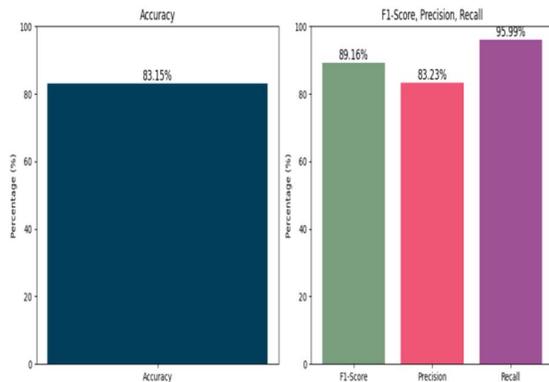


Figure 6. Evaluation Graph

Pada grafik kiri ditampilkan nilai akurasi model, sedangkan grafik kanan menunjukkan nilai F1-score, precision, dan recall. Dari hasil evaluasi yang telah dijelaskan, dapat disimpulkan bahwa:

1. Akurasi sebesar 83.15% menunjukkan bahwa model sudah cukup baik dalam memprediksi komentar bot dan komentar manusia secara keseluruhan.
2. Precision sebesar 83.23% berarti bahwa dari seluruh komentar yang diprediksi sebagai bot, sebagian besar memang benar bot. Ini penting agar komentar manusia tidak salah ditandai sebagai bot.
3. Recall sebesar 95.99% menunjukkan bahwa model sangat baik dalam menemukan komentar bot yang sebenarnya.
4. F1-score sebesar 89.16% menandakan bahwa model memiliki keseimbangan yang baik antara precision dan recall.

Kesimpulannya, model *Gradient Boosting* cukup efektif dalam mendeteksi komentar bot di Shopee. Walaupun masih terdapat sejumlah kecil komentar manusia yang salah diklasifikasikan sebagai bot (false positive), performa model secara keseluruhan sudah dapat dikatakan memuaskan. Untuk peningkatan lebih lanjut, model dapat ditingkatkan dengan penambahan data latihan, pemilihan fitur yang lebih spesifik, atau pengaturan parameter yang lebih optimal.

C. Deployment

Sistem ini dibangun menggunakan framework Flask sebagai backend untuk mengelola proses prediksi, dan antarmuka pengguna dikembangkan menggunakan HTML, CSS, serta sedikit JavaScript untuk memberikan tampilan yang interaktif dan modern.

Pada Figure 7 ditampilkan tampilan awal dari aplikasi web "*Deteksi Komentar Bot*" antarmuka aplikasi ini dirancang secara sederhana dan ramah pengguna (*user friendly*) agar mudah dioperasikan oleh siapa saja. Setelah mengetik komentar, pengguna cukup menekan tombol "Deteksi Sekarang" untuk memulai proses analisis. Komentar yang dimasukkan akan dikirim ke backend dan diproses oleh model Gradient Boosting yang telah dilatih sebelumnya. Model ini akan menganalisis pola dalam komentar, seperti struktur kalimat, pengulangan kata, atau penggunaan kata-kata hiperbolik yang sering muncul pada komentar otomatis dari bot. Hasil klasifikasi ditampilkan secara langsung di bawah tombol deteksi dalam bentuk kotak hasil, yang menunjukkan apakah komentar termasuk bot atau bukan bot. Tampilan ini dirancang agar proses deteksi cepat, praktis, dan mudah dipahami oleh pengguna.

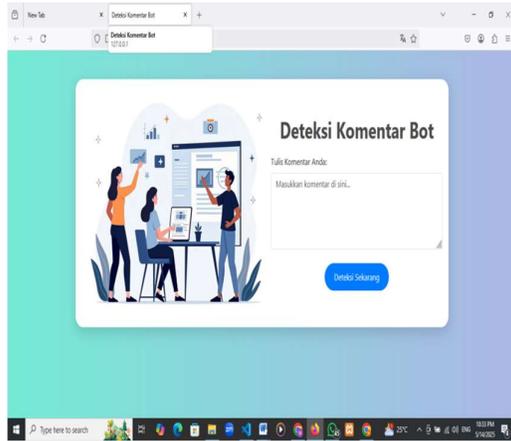


Figure 7. Website User Interface

Figure 8 menunjukkan hasil deteksi ketika pengguna memasukkan komentar yang sangat panjang dan terstruktur secara berlebihan, seperti: “*Produk ini benar-benar luar biasa, kualitasnya sangat bagus, sesuai dengan deskripsi, bahan sangat nyaman dipakai, pengiriman cepat dan aman...*”. Berdasarkan analisis model, komentar tersebut diklasifikasikan sebagai “Komentar Bot”, karena mengandung banyak indikator khas bot seperti struktur yang terlalu sempurna, penggunaan kata positif secara berlebihan, dan pola yang cenderung tidak natural untuk manusia.

Deteksi Komentar Bot



Figure 8. Detection Result

Sebaliknya, pada Figure 9 ditampilkan hasil deteksi untuk komentar yang lebih natural dan realistis seperti “*Saya beli ini karena ada promo, ternyata lumayan juga kualitasnya buat harian.*” Komentar ini dikategorikan sebagai “Bukan Bot”, karena bahasanya lebih santai, tidak terlalu repetitif, dan mencerminkan pengalaman subjektif yang lebih otentik dari pengguna manusia.

Deteksi Komentar Bot

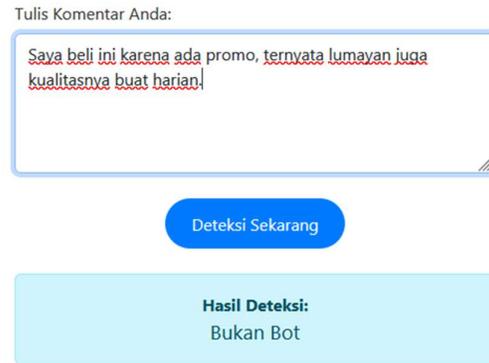


Figure 9. Detection Result

V. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan Dengan hadirnya sistem deteksi komentar bot yang telah dibuat, proses penyaringan komentar menjadi jauh lebih mudah dan efisien, terutama bagi pengguna non-teknis seperti pemilik toko online atau admin platform e-commerce. Mereka tidak perlu memiliki pengetahuan mendalam tentang *machine learning* untuk dapat menggunakannya. Cukup dengan mengakses aplikasi web, pengguna sudah bisa menguji komentar secara langsung tanpa instalasi tambahan. Hal ini membuktikan bahwa integrasi model *machine learning* ke dalam antarmuka web dapat menghasilkan solusi praktis, cepat, dan aplikatif yang siap digunakan dalam situasi nyata, mendukung kebutuhan bisnis secara langsung dan efisien.

Bahwa sistem deteksi komentar bot pada platform Shopee yang dibangun menggunakan algoritma *Gradient Boosting* mampu membedakan komentar bot dan manusia dengan cukup baik. Komentar bot berhasil diidentifikasi berdasarkan pola linguistik tertentu seperti struktur kalimat kaku, penggunaan pujian berulang, dan kurangnya konteks personal. Model yang diimplementasikan menunjukkan performa yang baik dengan akurasi 83.15% dan F1-score 89.16%, serta telah berhasil diterapkan dalam bentuk aplikasi web berbasis Flask yang mudah digunakan oleh admin toko. Meskipun demikian, masih terdapat tantangan seperti false positive yang memerlukan penyempurnaan lebih lanjut. Oleh karena itu, disarankan untuk menambah dan mendiversifikasi dataset, menerapkan teknik feature engineering yang lebih kaya (seperti analisis sentimen dan deteksi emoji), serta membandingkan algoritma *Gradient Boosting* dengan metode lain seperti SVM atau model deep learning. Pengembangan sistem ke arah dashboard interaktif dan integrasi pada skala nyata seperti Shopee Seller Center juga direkomendasikan agar sistem dapat memberikan dampak yang lebih luas dalam meningkatkan kepercayaan konsumen terhadap ulasan produk.

REFERENSI

DwiYansaputra, Ramaditia, Gibran Satya Nugraha, Fitri Bimantoro, and Arik Aranta. *Deteksi Sms Spam Berbahasa Indonesia Menggunakan Tf-Idf Dan*

- Stochastic Gradient Descent Classifier (Indonesian SMS Spam Detection Using TF-IDF and Stochastic Gradient Descent Classifier)*. <http://jtika.if.unram.ac.id/index.php/JTIKA/>.
- Elsa Suryana, Silvia, and Budi Warsito. "Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank." 10: 617–23. <https://ejournal3.undip.ac.id/index.php/gaussian/>.
- Gudiato, C, E Sedyono, and I Sembiring. 2022. "Analisis Sistem E-Commerce Pada Shopee Untuk Meningkatkan Daya Saing Menggunakan Metode S.W.O.T." *JIFOTECH (Journal of Information Technology)* 2(1): 6–9. doi:10.46229/jifotech.v2i1.294.
- Kurniawan, Hendra, and Maritim Raja Ali Haji Jl Politeknik Senggarang. 2020. "Jurnal Sustainable: Jurnal Hasil Penelitian Dan Industri Terapan Deteksi Twitter Bot Menggunakan Klasifikasi Decision Tree." 09(01): 31–37.
- Lestari, Tri Putri. 2022. "Analisis Text Mining Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) Dan Social Network Analysis (SNA)." *Jurnal Informatika Ekonomi Bisnis*: 65–71. doi:10.37034/infec.v4i3.146.
- Muktafin, E H, K Kusriani, and E T Luthfi. 2020. "Analisis Sentimen Pada Ulasan Pembelian Produk Di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing." *Jurnal Eksplorasi Informatika* 10(1): 32–42. doi:10.30864/eksplorasi.v10i1.390.
- Muktafin, Elik Hari, Kusriani Kusriani, and Emha Taufiq Luthfi. 2020. "Analisis Sentimen Pada Ulasan Pembelian Produk Di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing." *Jurnal Eksplorasi Informatika* 10(1): 32–42. doi:10.30864/eksplorasi.v10i1.390.
- Munawar, Zen, Herru Soerjono, Novianti Indah Putri, Hernawati, and Andina Dwijayanti. 2023. "Manfaat Kecerdasan Buatan ChatGPT Untuk Membantu Penulisan Ilmiah." *TEMATIK* 10(1): 54–60. doi:10.38204/tematik.v10i1.1291.
- Nurdin, Rizal, and Rizwan. 2019. 1 *Jurnal Telematika Pendeteksian Dokumen Plagiarisme Dengan Menggunakan Metode Weight Tree*.
- Priyatno, A M, M M Muttaqi, F Syuhada, and A Z Arifin. 2019. "Deteksi Bot Spammer Twitter Berbasis Time Interval Entropy Dan Global Vectors for Word Representations Tweet's Hashtag." *Register: Jurnal Ilmiah Teknologi Sistem Informasi* 5(1): 37–46. doi:10.26594/register.v5i1.1382.
- Puspitasari, Agreianti, Astrid Noviana Paradhita, Yohanes Wien Tineka, Vivin Sulistyowati, Ni Komang Septia Noriska, and Haryanto. 2024. "Natural Language Processing (NLP) Technology for Chatbot Website." *Jurnal Penelitian Pendidikan IPA* 10(SpecialIssue): 319–24. doi:10.29303/jppipa.v10ispecialissue.8241.
- Rayadin, M Amhar, M Musaruddin, and R Adi Saputra. 2024. "Implementasi Ensemble Learning Metode XGBoost Dan Random Forest Untuk Prediksi Waktu Penggantian Baterai Aki." *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer* 5(2): 111–19. doi:10.37148/bios.v5i2.128.
- Rayadin, Muhamad Amhar, Mustarum Musaruddin, Rizal Adi Saputra, and Isnawaty Isnawaty. 2024a. "Implementasi Ensemble Learning Metode XGBoost Dan Random Forest Untuk Prediksi Waktu Penggantian Baterai Aki." *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer* 5(2): 111–19. doi:10.37148/bios.v5i2.128.
- Rayadin, Muhamad Amhar, Mustarum Musaruddin, Rizal Adi Saputra, and Isnawaty Isnawaty. 2024b. "Implementasi Ensemble Learning Metode XGBoost Dan Random Forest Untuk Prediksi Waktu Penggantian Baterai Aki." *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer* 5(2): 111–19. doi:10.37148/bios.v5i2.128.
- Ridwansyah, Muhammad, and Hadi Zakaria. 2023. 1 *JURIHUM: Jurnal Inovasi dan Humaniora Implementasi Algoritma Gradient Boosting Pada Aplikasi Hutang Piutang Perorangan Secara Berbasis Web Untuk Meningkatkan Akurasi Prediksi Pelunasan Hutang (Studi Kasus: PT Naila Kreasi Mandiri)*. <https://jurnalmahasiswa.com/index.php/jurikum>.
- Rosnita, Lidya. 14 *Dengan Optical Character Recognition Di Orbit Future Academy*.
- Rosnita, Lidya, Sujacka Retno, and Suhaiba Nasyira Hariono. 15 *PENERAPAN SISTEM DETEKSI PENGISIAN RUANG PARKIR KENDARAAN RODA 4 MENGGUNAKAN METODE COMPUTER VISION DI ORBIT FUTURE ACADEMY*.
- Rusmarasy, Bestralaga, Bayu Priyambadha, and Fajar Pradana. 2019. 3 *Pengembangan Chat Bot Pada CoMa Untuk Memberikan Motivasi Kepada Pengguna Menggunakan AIML*. <http://j-ptiik.ub.ac.id>.
- Urlamma, D., M. Supriya, D. Lavanya, and A. Hari Priya. 2024. "Detection Of Phishing Websites Using Gradient Boosting Classifier Based On URL." *IARJSET* 11(3). doi:10.17148/iarjset.2024.11318.
- Urlamma, D, M Supriya, D Lavanya, and A Hari Priya. 2024. "Detection Of Phishing Websites Using Gradient Boosting Classifier Based On URL." *International Advanced Research Journal in Science, Engineering and Technology* 11(3): 116–21. doi:10.17148/iarjset.2024.11318.
- Wijaya, Y F, S Yulianto, and J Prasetyo. 2021. "Model Penilaian Tata Guna Lahan Dengan Citra Landsat 8 OLI Menggunakan Algoritma XGBoost Diwilayah Beresiko Tsunami (Studi Kasus : Kota Palu Sulawesi Tengah)." *Indonesia Journal of Computing Modeling* 4(1): 24–28. doi:10.24246/icm.v4i1.4981.