

Pembelajaran *Ensemble* Untuk Klasifikasi Ulasan Pelanggan *E-commerce* Menggunakan Teknik *Boosting*

Matius Rama Hadi Suryanto^{1*}, Danang Wahyu Utomo²

^{1,2}Program Studi Teknik Informatika, Universitas Dian Nuswantoro

^{1,2}Jalan Imam Bonjol No. 207 Semarang Tengah, Semarang, Jawa Tengah, Indonesia

E-mail: 111202012395@mhs.dinus.ac.id¹, danang.wu@dsn.dinus.ac.id²

Abstrak

Info Naskah:

Naskah masuk: 22 Mei 2024

Direvisi: 25 Juni 2024

Diterima: 2 Juli 2024

Saat ini aktivitas jual beli telah banyak dilakukan pada layanan *e-commerce*. Tingginya pengguna *e-commerce* menjadi faktor utama dalam meningkatkan kualitas layanan *e-commerce*. Salah satu faktor peningkatan kualitas layanan *e-commerce* adalah ulasan pelanggan. Ulasan pelanggan berguna bagi pemilik toko untuk mengetahui produk yang ditawarkan memiliki ulasan positif atau negatif. Masalah utama adalah jumlah ulasan pelanggan yang banyak menjadi kendala pemilik toko dalam melakukan klasifikasi ulasan pelanggan. Penelitian ini mengusulkan klasifikasi ulasan pelanggan menggunakan pembelajaran ensemble dengan teknik boosting yaitu *XGBoost*, *AdaBoost*, *Gradient Boosting* dan *LightGBM*. Teknik yang diusulkan ditujukan untuk meningkatkan performa algoritma. Pada skenario uji, teknik *majority voting* digunakan untuk menghasilkan performa terbaik dari masing – masing algoritma. Hasil menunjukkan bahwa algoritma *XGBoost* menghasilkan akurasi lebih tinggi dari teknik boosting lainnya yaitu sebesar 92.30%. Pada analisis matriks evaluasi presisi, recall, dan F1-Score *XGBoost* menghasilkan nilai true positif lebih tinggi dibandingkan Teknik lainnya seperti *AdaBoost*, *Gradient Boosting* dan *Light GBM*.

Abstract

Keywords:

classification;
sentiment analysis;
e-commerce;
ensemble learning;
boosting.

Selling and buying activities are carried out in e-commerce services. The high number of e-commerce users is the main factor in improving the quality of e-commerce services. One factor in improving the quality of e-commerce services is customer reviews. Customer reviews are useful for shop owners to find out whether the product has positive or negative reviews. The main problem is that the large number of customer reviews is an obstacles for shop owners in classifying customer reviews. This study proposes classifying customer reviews using ensemble learning with boosting techniques such as XGBoost, AdaBoost, Gradient Boosting, and LightGBM. The proposed technique is aimed at improving the performance of the algorithm. In a test scenario, majority voting use to produce the best performance from each algorithm. The result shows that the XGBoost algorithm produces higher accuracy than other techniques are 92.30%. In matic evaluation of precision, recall, and F1-Score, XGBoost produces higher true positive values than other techniques such as AdaBoost, Gradient Boosting, and Light GBM.

*Penulis korespondensi:

Matius Rama Hadi Suryanto

E-mail: 1111202012395@mhs.dinus.ac.id

1. Pendahuluan

Perkembangan teknologi telah berkembang pesat dan memberikan dampak perubahan perilaku dalam kehidupan sehari-hari. Salah satu perubahan besar adalah aktivitas jual beli. Saat ini aktivitas jual beli banyak dilakukan melalui online. Melalui *e-commerce*, platform yang memberikan kemudahan aktivitas jual beli bagi pengguna. Menurut [1], *e-commerce* memberikan kemudahan bagi para pelaku bisnis dalam memasarkan produknya secara luas. Pemasaran produk dapat disebarluaskan secara luas untuk mendapatkan pembeli atau konsumen yang lebih besar. Adanya *e-commerce* terbukti menjadi faktor pendorong utama pelanggan menggunakan platform *e-commerce* [2]. Dari sisi pelanggan, *e-commerce* memberikan kemudahan dalam pencarian produk tanpa terbatas ruang dan waktu.

Peningkatan *e-commerce* ditandai dengan tingginya jumlah pengguna pada platform jual beli online. Menurut laman website dataindonesia.id jumlah pengguna *e-commerce* di Indonesia pada tahun terakhir 2023 adalah 196,47 juta jiwa. Transaksi jual beli dalam platform *e-commerce* melibatkan data pengguna dan produk. Dapat disimpulkan bahwa banyaknya pengguna *e-commerce* maka data produk dan data pengguna yang digunakan juga besar. Layanan dalam platform menyediakan produk dalam jumlah besar [3]. Produk yang ditawarkan dalam platform *e-commerce* memiliki bermacam-macam variasi jumlah dan deskripsi produk. [4].

Peningkatan layanan *e-commerce* terus dilakukan untuk memberikan layanan terbaik kepada para pelanggan. Peningkatan layanan tidak hanya fokus pada kemudahan dalam transaksi namun juga fokus pada sisi kemudahan pelanggan dalam mengoperasikan perangkat, sistem memberikan rekomendasi produk, dan kategorisasi produk. Selain itu, peningkatan layanan *e-commerce* dan produk juga dapat melibatkan pengguna yaitu melalui ulasan pelanggan. Beberapa peneliti telah mengusulkan alat, metode, dan model dalam meningkatkan layanan *e-commerce*: penggunaan kecerdasan buatan dalam *e-commerce* [5], klasifikasi produk [6], kategorisasi produk [7], analisis sentimen ulasan pelanggan [8], analisis sentimen ulasan produk [9].

Ulasan pelanggan menjadi satu hal yang penting dalam upaya peningkatan kualitas layanan *e-commerce*. Ulasan pelanggan adalah hasil review atau komentar dari pelanggan terhadap produk yang ditawarkan. Ulasan pelanggan sangat berguna bagi pelaku bisnis untuk mengetahui kualitas dari produk dari sisi konsumen [10]. Ulasan atau komentar memberikan beragam informasi dan umpan balik terhadap kualitas produk yang ditawarkan. Komentar berupa kritik, saran dan rating dari angka dapat menjadi evaluasi dalam peningkatan layanan *e-commerce*. Ulasan juga digunakan untuk mengetahui tren dari produk yang banyak diminati [11]. Dari beragam ulasan tersebut dapat dianalisis untuk mengetahui produk mana yang banyak diminati pelanggan dan mengikuti tren saat ini.

Ulasan pelanggan menjadi bagian penting bagi pelaku bisnis dalam meningkatkan kualitas layanan *e-commerce*. Namun, permasalahan juga ditemukan dalam penggunaan ulasan pelanggan. Beragamnya ulasan pelanggan dan banyaknya produk yang diulas menjadi faktor pelaku bisnis

atau pemilik toko kesulitan dalam melakukan analisis ulasan pelanggan [12]. Pemilik toko membutuhkan waktu untuk melakukan analisis ulasan pelanggan guna mengetahui ulasan yang baik atau sebaliknya. Menurut Istiqamah [13], ketidakseimbangan jumlah ulasan pelanggan juga menjadi masalah dalam menganalisis ulasan pelanggan. Ulasan yang tidak seimbang dapat menjadikan kesalahan dalam menganalisis ulasan pelanggan. Jika diterapkan pada model atau algoritma dapat menjadikan performa tidak optimal.

Berdasarkan penelitian terdahulu, solusi terkait permasalahan analisis ulasan pengguna telah diusulkan. Pendekatan, model dan algoritma diusulkan untuk mengatasi permasalahan ulasan pelanggan: algoritma *XGBoost* [14], *ensemble learning* [15], *random forest* [16], *support vector machine* (SVM) [17], *K Nearest Neighbor* (KNN) [18]. Solusi ulasan pelanggan berbasis klasifikasi teks dianggap mampu menangani permasalahan dalam analisis sentimen ulasan pelanggan.

Berdasarkan uraian diatas, studi ini mengusulkan pembelajaran ensemble (*ensemble learning*) untuk klasifikasi ulasan pelanggan. Pembelajaran ensemble menggabungkan algoritma dengan performa yang lemah, kemudian dilatih dengan model baru untuk mendapatkan performa yang lebih dari sebelumnya. Teknik *boosting* mampu memberikan akurasi lebih baik dibandingkan algoritma individu dalam pembelajaran mesin [19]. Pada penelitian ini menggunakan algoritma *XGBoost* dengan pendekatan teknik *boosting* yang digunakan untuk menggabungkan beberapa pohon keputusan (*decision tree*) untuk meningkatkan performa dari model sebelumnya.

2. Metode

2.1 Ensemble Learning

Pembelajaran ensemble adalah teknik dalam pembelajaran mesin yang digunakan untuk meningkatkan kinerja dan akurasi model dengan menggabungkan hasil prediksi dari beberapa model atau algoritma yang berbeda. Teknik ini menggabungkan beberapa algoritma atau model dengan performa yang kurang baik untuk menghasilkan model dengan performa yang lebih baik dibandingkan pembelajaran tunggal. Menurut Mamun N [20], pembelajaran ensemble merupakan teknik yang menggabungkan pengklasifikasi dasar (*base classifier*) untuk mengembangkan model baru dengan menggunakan performa pengklasifikasi individu. Teknik ensemble menghitung komputasi rata-rata probabilitas dari pengklasifikasi dasar. Hasil maksimum dari probabilitas tersebut digunakan sebagai label akhir.

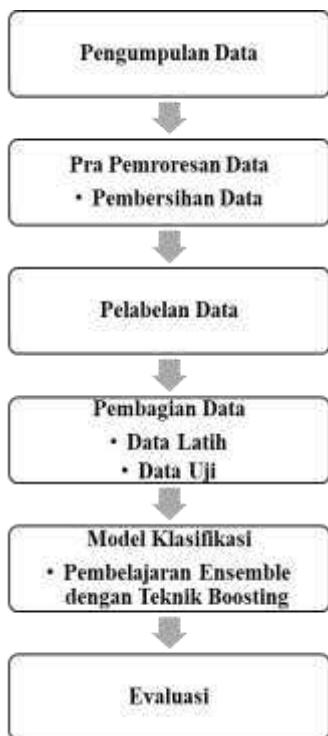
Pada penelitian ini pembelajaran ensemble diusulkan untuk menggabungkan teknik *boosting* dengan tujuan mencari performa terbaik dari algoritma tunggal seperti *XGBoost*, *AdaBoost*, *Gradient Boosting* dan *LightGBM* tunggal. Dari algoritma tunggal tersebut dijadikan sebagai pengklasifikasi dasar untuk mendapatkan hasil atau performa terbaik dari hasil pembelajaran ensemble

2.2 Teknik Boosting

Teknik *boosting* adalah teknik dalam pembelajaran mesin yang melatih pembelajar (*learner*) dengan performa yang lemah untuk menghasilkan model dengan performa

yang kuat. Secara teknis, teknik *boosting* melatih pembelajar lemah, menghitung prediksi, memilih sampel yang terkласifikasi tidak sesuai, melatih pembelajar lainnya dengan sampel salah dari iterasi sebelumnya. Salah satu tujuan utama teknik *boosting* adalah meningkatkan pembelajar lemah dengan bias yang tinggi [21]. Algoritma seperti *AdaBoost*, *XGBoost* memberikan performa yang lebih baik. Menurut Bertini Junior, metode ensemble dengan teknik *boosting* mampu diterapkan dalam pembelajaran mesin dan memberikan hasil terbaik [22].

Pada penelitian ini menggunakan pembelajaran ensemble dengan teknik *boosting* seperti *AdaBoost*, *XGBoost*, *Gradient Boosting* dan *LightGBM*. Tujuannya, untuk menguji pembelajar dengan performa yang lemah dan mendapatkan hasil dari pembelajaran dengan akurasi lebih baik dari algoritma individu.



Gambar 1. Alur eksperimen

2.3 Eksperimen

Berdasarkan alur eksperimen pada Gambar 1, tahap pertama dimulai dari pengumpulan data yaitu menggunakan data publik *flipkart* yang diambil dari laman <https://www.kaggle.com/datasets/kabirnagpal/flipkart-customer-review-and-rating>. Dataset tersebut merupakan kumpulan ulasan pengguna yang fokus pada produk *headphone*. Dataset terdiri dari 2 kolom yaitu *review* dan *rating* dengan jumlah 9976. Format dataset adalah *csv*.

Tahap kedua adalah pra pemrosesan data yaitu tahap pembersihan data dengan cara mengubah ulasan ke konversi huruf kecil, menghilangkan kata – kata tidak relevan yang sering muncul (*stopword removal*), mengganti kata dengan pola kata yang baku (*re sub*), mengecek dan menghapus data yang kosong, dan data ganda. Sebagai contoh kata pada dataset baris ke 1 yaitu

Kata sebelum pra pemrosesan:

"It was nice produt. I like it's design a lot. It's easy to carry. And. Looked stylish."

Kata setelah pra pemrosesan:

"nice produt like design lot easy carry looked stylish"

pada contoh diatas, beberapa kata telah dihilangkan setelah dilakukan proses pra pemrosesan seperti kata 'I', 'it's', 'it was'.

Tahap ketiga adalah pelabelan data yaitu memberikan label data seperti positif (1) dan negative (0) untuk memudahkan dalam komputasi probabilitas label. Tahap keempat adalah pembagian data (*split data*) yaitu dataset dibagi kedalam data uji dan data latih untuk diterapkan pada model klasifikasi. Tahap kelima pengaturan model klasifikasi yaitu pengaturan *XGBoost* dalam pembelajaran ensemble dengan menggunakan *majority voting*. Tahap keenam adalah evaluasi menggunakan *confusion matrix* untuk mengetahui model dengan membandingkan prediksi model dari nilai aktual data. Tabel *confusion matrix* menghasilkan nilai *true positive*, *true negative*, *false positive*, *false negative*. Dari matriks ini dapat digunakan untuk mengetahui nilai *akurasi*, *presisi*, *recall* dan *f1-score*.

2.4 Skenario Uji

Skenario uji dimulai dengan pengaturan data uji dan data latih yaitu data uji 20% dan data latih 80%. Pengaturan selanjutnya adalah pengaturan parameter pada pelatihan *XGBoost* atau biasa disebut pengaturan *hyperparameter* yang meliputi: *n_estimator*, *learning_rate* dan *max_depth*. Berikut Tabel I daftar pengaturan hyperparameter pada *XGBoost*

Tabel 1 Hyperparameter Tuning

| Model | N_estimator | Learning_rate | Max_depth |
|---|-------------|---------------|-----------|
| XGBoost, Gradient Boosting, LightGBM | 100 | 0.1 | [3, 5, 7] |
| | 200 | 0.2 | [3, 5, 7] |
| <i>AdaBoost</i> | 100 | 0.1 | - |
| | 200 | 0.2 | 0.2 |

Berdasarkan pengaturan parameter pada Tabel I dengan teknik pencarian acak (*random search*) digunakan untuk mengetahui perbandingan performa dari masing – masing paramter. Khusus algoritma *AdaBoost* pengaturan hyperparameter hanya menggunakan *n_estimator* dan *learning_rate*. Algoritma *AdaBoost* tidak menggunakan parameter *max_depth*.

3. Hasil dan Pembahasan

3.1 Hasil Pra-Pemrosesan Data

Eksperimen dimulai dengan prapemrosesan data (*preprocessing data*) yaitu pembersihan data. Data dinormalisasi untuk dihilangkan beberapa hal seperti: ulasan ganda (redundant data) dan ulasan kosong. Selain itu, pembersihan juga dilakukan untuk tiap – tiap ulasan seperti penghilangan tanda baca, penghapusan kata tidak relevan

(stopword removal), pengubahan kata dalam bentuk lowercase, dan mengubah kata menjadi kata dasar. Prapemrosesan data ditujukan untuk memudahkan penghitungan pada model algoritma *XGBoost*.

| | review | rating | negative | neutral | positive |
|---|---|--------|----------|---------|----------|
| 0 | It was nice product. I like its design a lot, ... | 5 | 0.0 | 0.513 | 0.387 |
| 1 | awesome sound...very pretty to see this rd th... | 5 | 0.0 | 0.517 | 0.483 |
| 2 | awesome sound quality. pros 7-8 hrs of battery... | 4 | 0.0 | 0.823 | 0.177 |
| 3 | I think it is such a good product not only as ... | 5 | 0.0 | 0.672 | 0.328 |
| 4 | awesome bass sound quality very good battery i... | 5 | 0.0 | 0.586 | 0.405 |

Gambar 2. Sample Hasil SentimentIntensity Analyze

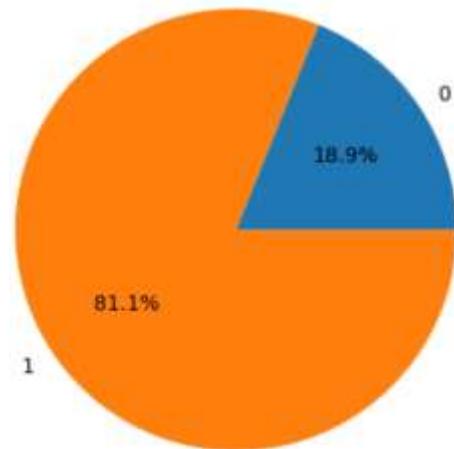
Prapemrosesan selanjutnya adalah pelabelan data, dari parameter review dan rating diberikan kolom baru yang *class label*, yang ditujukan sebagai penentu ulasan masuk dalam kategori positif atau negatif. Pelabelan dengan cara memberikan nilai 1 untuk ulasan yang dianggap positif dan nilai 0 untuk ulasan yang dianggap negatif. Pada implementasi menggunakan library *SentimentIntensityAnalyzer*. Library tersebut menampilkan hasil masing - masing ulasan mendapatkan nilai positif, negative dan netral. Nilai yang ditampilkan dalam rentang 0-1. Pada Gambar 1 menunjukkan bahwa masing - masing ulasan memiliki nilai positif, negatif, dan netral. Sebagai contoh, pada ulasan indeks ke 0 menunjukkan nilai positive 0.367 maka dapat diartikan bahwa ulasan tersebut adalah sentiment positif.

| | review | rating | negative | neutral | positive | compound | class label |
|---|---|--------|----------|---------|----------|----------|-------------|
| 0 | It was nice product. I like its design a lot, ... | 5 | 0.0 | 0.613 | 0.387 | 0.0020 | 1 |
| 1 | awesome sound...very pretty to see this rd th... | 5 | 0.0 | 0.517 | 0.483 | 0.0001 | 1 |
| 2 | awesome sound quality. pros 7-8 hrs of battery... | 4 | 0.0 | 0.823 | 0.177 | 0.0019 | 1 |
| 3 | I think it is such a good product not only as ... | 5 | 0.0 | 0.672 | 0.328 | 0.0001 | 1 |
| 4 | awesome bass sound quality very good battery i... | 5 | 0.0 | 0.586 | 0.405 | 0.0053 | 1 |

Gambar 3. Sampel Hasil Pelabelan ‘class label’

Selanjutnya pemberian label didasarkan pada polaritas label ‘rating’. Hal ini digunakan untuk mengetahui dari ulasan dan rating yang digunakan apakah produk memiliki sentiment positif atau negatif. Pada Gambar 2 yaitu pada label ‘class label’ berisi nilai 1 dan 0. Nilai 1 didasarkan pada rating diatas atau sama dengan 4 sedangkan nilai 0 didasarkan pada rating dibawah 4.

Hasil dari pelabelan tersebut ditampilkan dalam bentuk grafik untuk mengetahui apakah dataset yang digunakan termasuk dataset imbalanced atau balanced. Berdasarkan Gambar 4, visualisasi ditampilkan dalam bentuk diagram pie, menunjukkan bahwa jumlah pelabelan untuk kelas 1 81.1% dan kelas 0 adalah 18.9%. dari visualisasi ini menunjukkan bahwa dataset dikategorikan dataset imbalanced karena adanya ketidakseimbangan dalam jumlah label yang dihasilkan. Prosentase yang dihasilkan kelas 1 lebih tinggi dibandingkan kelas 0.



Gambar 4. Visualisasi Jumlah Class Label 1 dan 0

Perlu dilakukan penyeimbangan dataset atau pengaturan agar dataset seimbang (*balance*). *Upsampling* adalah salah satu metode yang digunakan untuk penambahan manual pada kelas minoritas agar jumlah antara kelas sebanding. Dalam hal ini jumlah kelas 0 dan kelas 1 sama atau prosentase yang dihasilkan 50%. Tujuan utama dari *upsampling* adalah untuk keseimbangan data, meningkatkan performa model yang diusulkan dan mencegah terjadinya overfitting atau prediksi yang tidak akurat untuk data baru.

3.2 Pembagian data

Setelah dilakukan prapemrosesan data, selanjutnya adalah pembagian data yaitu dibagi ke dalam data latih dan data uji. Ukuran data uji yang digunakan adalah 0.2 atau 20% dari total jumlah dataset yang digunakan. Jadi, jumlah data latih yang digunakan adalah 8091 dan data uji yang digunakan adalah 1885.

3.3 Hasil Pemodelan

Pada uji eksperimen menggunakan pembelajaran *ensemble* dengan teknik boosting yaitu algoritma *AdaBoost*, *XGBoost*, *Gradient Boosting* dan *LightGBM*. Masing - masing algoritma dijalankan dengan pengaturan hyperparameter seperti pada Tabel I. implementasi pembelajaran ensemble menggunakan Teknik *majority voting* untuk mendapatkan nilai terbaik dari pembelajar. Estimator pada Teknik voting menggunakan Pembelajar dasar dalam hal ini adalah masing - masing algoritma. Tujuannya, untuk melatih algoritma individu dengan performa lemah untuk mendapatkan hasil performa yang lebih baik.

Pada Tabel 2 menunjukkan hasil akurasi dengan pengaturan parameter yang telah ditentukan. Uji pertama adalah pembelajar *XGBoost*. Hasil terbaik uji Ensemble *XGBoost* pada *n_estimator* 200, *learning rate* 0.2 dan *max depth* 9 dengan akurasi 92.30%. dari semua hasil uji *XGBoost*, tingginya *n_estimator*, *learning rate* dan *max depth* mempengaruhi hasil akurasi *XGBoost*. Semakin besar nilai pada hyperparameter, nilai akurasi yang dihasilkan lebih besar.

Tabel 2 Hasil Pengujian Model
Hyperparameter

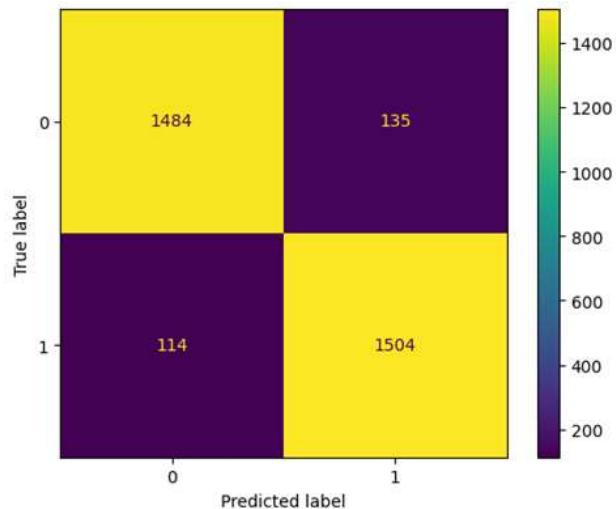
| Model | <i>n_estimator</i> | <i>learning rate</i> | <i>max depth</i> | Hasil |
|-------------------------------------|--------------------|----------------------|------------------|---------------|
| XGBoost | 100 | 0.1 | 5 | 87.17% |
| | | | 7 | 88.63% |
| | | | 9 | 89.77% |
| | 200 | 0.2 | 5 | 88.69% |
| | | | 7 | 90.42% |
| | | | 9 | 91.28% |
| | 100 | 0.1 | 5 | 88.94% |
| | | | 7 | 90.05% |
| | | | 9 | 91.19% |
| AdaBoost | 200 | 0.2 | 5 | 90.67% |
| | | | 7 | 91.78% |
| | 100 | 0.2 | 9 | 92.30% |
| | | | - | |
| Gradient Boosting (GradBoost) | 100 | 0.1 | 5 | 86.83% |
| | | | 7 | 88.81% |
| | | | 9 | 90.57% |
| | 200 | 0.2 | 5 | 89.15% |
| | | | 7 | 90.73% |
| | | | 9 | 91.25% |
| | 100 | 0.1 | 5 | 89.06% |
| | | | 7 | 90.88% |
| | | | 9 | 91.62% |
| LightGBM | 200 | 0.2 | 5 | 90.73% |
| | | | 7 | 91.38% |
| | | | 9 | 91.68% |
| | 100 | 0.1 | 5 | 85.85% |
| | | | 7 | 87.21% |
| | | | 9 | 87.79% |
| | 200 | 0.2 | 5 | 87.61% |
| | | | 7 | 89.09% |
| | | | 9 | 89.86% |
| | 100 | 0.1 | 5 | 87.95% |
| | | | 7 | 89.31% |
| | | | 9 | 89.83% |
| | 200 | 0.2 | 5 | 89.65% |
| | | | 7 | 90.36% |
| | | | 9 | 91.56% |

Uji kedua adalah pembelajar *AdaBoost*. Pada pembelajar *AdaBoost* hanya menggunakan parameter *n_estimator* dan *learning rate*. Hasil akurasi terbaik adalah 83.34% pada parameter *n_estimator* 200 dan *learning rate* 0.2. Tingginya *n estimator* dapat meningkatkan akurasi pada *AdaBoost*. Terbukti, dengan *learning rate* yang sama (misal 0.1) namun *n_estimator* berbeda, nilai *n_estimator* 200 menghasilkan akurasi lebih tinggi dibandingkan *n_estimator* 100. Uji ketiga adalah pembelajar Gradient Boosting. Uji model menggunakan pengaturan hyperparameter yang sama dengan uji *XGBoost*. Hasil akurasi terbaik pada pembelajar *Gradient Boosting* adalah 91.68% pada pengaturan *n_estimator* 200, *learning rate* 0.2 dan *max_depth* 9. Sama seperti pembelajar *XGBoost*, pada *Gradient Boosting* akurasi dapat dipengaruhi oleh *n_estimator*, *learning rate*, dan *max depth*. Semakin tinggi nilai pada parameter tersebut semakin tinggi nilai akurasi yang dihasilkan.

Uji keempat adalah pembelajar LightGBM. Hasil akurasi terbaik dihasilkan pada *n estimator* 200, *learning rate* 0.2, dan *max depth* 9 yaitu dengan nilai akurasi 91.56. Dari hasil ini, faktor peningkatan performa pembelajar LightGBM juga dipengaruhi oleh parameter *n_estimator*, *learning rate*, dan *max depth*.

Dari keseluruhan hasil uji dapat ditarik kesimpulan bahwa pembelajar *XGBoost* menghasilkan performa terbaik dibandingkan pembelajar *AdaBoost*, *Gradient Boosting* dan LightGBM. Dalam pengaturan hyperparameter yang sama antara *XGBoost*, *Gradient Boosting* dan LightGBM, pembelajar *XGBoost* unggul pada masing – masing hyperparameter. Sebagai contoh, pada *n_estimator* 100, *learning rate* 0.1, dan *max_depth* 5 pembelajar *XGBoost* menghasilkan akurasi 87.17%, *Gradient Boosting* 86.83% dan LightGBM 85.85%. Dalam pengaturan nilai parameter terendah, pembelajar *XGBoost* tetap unggul melebihi pembelajar *Gradient Boosting* dan LightGBM

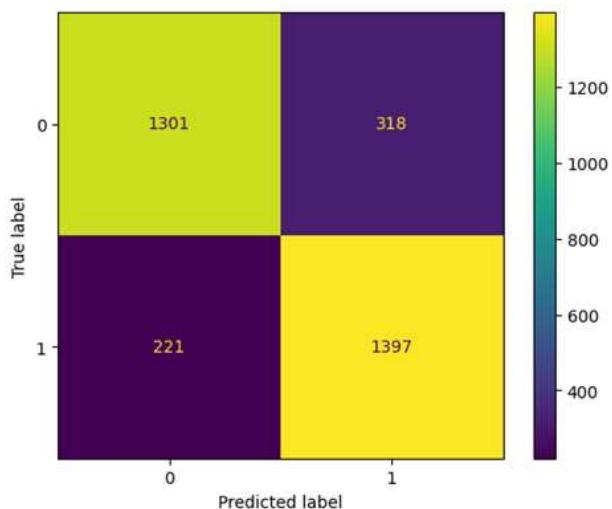
Analisis pengujian model ditampilkan dalam matriks yaitu *confusion matrix*. Confusion matrix adalah evaluasi performa dalam bentuk tabel yang membandingkan model yang diprediksi dengan nilai aktual data. Pada uji pembelajar *XGBoost* dengan nilai akurasi tertinggi yaitu 92.30% maka *confusion matrix* ditampilkan seperti pada Gambar 5. *Confusion matrix* menunjukkan bahwa nilai *true positif* lebih besar dibandingkan nilai *false negative*, *false positif* dan *true negative* yaitu 1504.



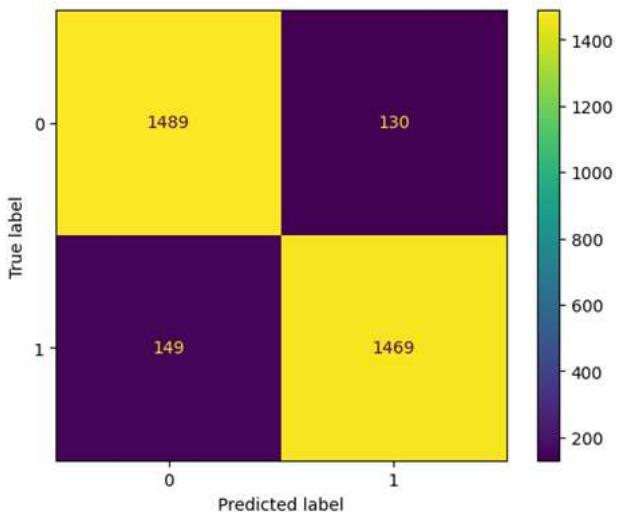
Gambar 5. Confusion Matrix - XGBoost

Pada Gambar 6 menunjukkan *Confusion Matrix* pada pembelajar *AdaBoost* dengan *n_estimator* 200 dan *learning rate* 0.2. Nilai kelas *true positive* lebih besar yaitu 1397 dibandingkan kelas lainnya. Pada Gambar 7 menunjukkan *Confusion Matrix* pada pembelajar *Gradient Boosting* dengan parameter *n_estimator* 200, *learning rate* 0.2 dan *max_depth* 9. Pada *Gradient Boosting* nilai kelas terbesar adalah *true negative* dengan nilai 1489.

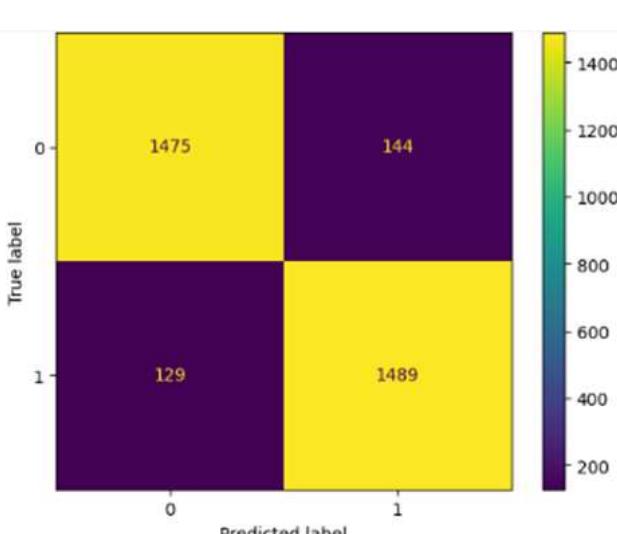
Pada Gambar 8, menunjukkan bahwa tabel *Confusion Matrix* LightGBM menghasilkan *true positive* 1489 pada *n_estimator* 200, *learning rate* 0.2 dan *max depth* 9. Nilai ini lebih tinggi dibandingkan kelas lainnya.



Gambar 6. Confusion Matrix – AdaBoost



Gambar 7. Confusion Matrix – Gradient Boosting



Gambar 8. Confusion Matrix - LightGBM

Berdasarkan hasil yang ditampilkan pada *confusion matrix* ketiga algoritma yaitu *XGBoost*, *AdaBoost* dan *LightGBM* menghasilkan nilai *true positive* lebih besar dibandingkan kelas lain. Hal ini membuktikan bahwa pembelajaran atau model yang digunakan mampu menghasilkan prediksi yang sesuai dengan aktual. Selaras dengan nilai akurasi yang dihasilkan, algoritma *XGBoost* menghasilkan nilai *true positive* lebih besar jika dibandingkan dengan *AdaBoost* dan *LightGBM*. Hal ini dapat disimpulkan bahwa algoritma *XGBoost* mampu memberikan peningkatan performa lebih baik setelah dilakukan pembelajaran ensemble. Sedangkan pada algoritma *Gradient Boosting* nilai *true positive* yang dihasilkan lebih kecil dibandingkan *true negative*. Perlu adanya peningkatan pada algoritma *Gradient Boosting* untuk meningkatkan nilai *true positive* seperti penanganan data tidak seimbang, penggunaan teknik selain *boosting* seperti *bagging* dan *stacking*.

Analisis evaluasi *confusion matrix* digunakan untuk mengetahui dengan pasti bahwa prediksi kelas positif benar – benar positif atau sesuai. Perlu adanya evaluasi melalui analisis presisi, recall dan f1-score. Evaluasi presisi digunakan untuk mengevaluasi seberapa akurat model mampu melakukan prediksi kelas positif yaitu benar – benar positif. Pada Tabel IV menunjukkan nilai *XGBoost* 0.92, *AdaBoost* 0.81, *Gradient Boosting* 0.92 dan *Light GBM* 0.91. Hanya algoritma *AdaBoost* yang menghasilkan presisi dibawah 0.90 atau 90%. Artinya, dari keempat Teknik boosting yang digunakan, hanya *AdaBoost* yang memiliki performa kurang baik dalam keakuratan menghasilkan prediksi kelas positif.

Evaluasi *recall* digunakan untuk mengevaluasi bahwa kelas positif yang diprediksi sesuai atau benar diidentifikasi. Hasil Tabel 3 menunjukkan bahwa algoritma *XGBoost*, *AdaBoost*, dan *Gradient Boosting* menghasilkan Recall diatas 90%. Artinya, algoritma tersebut mampu diimplementasikan pada klasifikasi ulasan pelanggan dan mampu menghasilkan prediksi klasifikasi yang benar.

Tabel 3 Analisis Matriks

| Model | Akurasi | Precision | Recall | F1-Score |
|------------------|---------|-----------|--------|----------|
| <i>XGBoost</i> | 92.30% | 0.92 | 0.93 | 0.92 |
| <i>AdaBoost</i> | 83.34% | 0.81 | 0.86 | 0.84 |
| <i>GradBoost</i> | 91.68% | 0.92 | 0.91 | 0.91 |
| <i>LightGBM</i> | 91.56% | 0.91 | 0.92 | 0.92 |

Evaluasi F1-score merupakan evaluasi keseimbangan antara presisi dan recall. Hasil Tabel IV menunjukkan bahwa *XGBoost*, *Gradient Boosting* dan *LightGBM* menghasilkan nilai diatas 90 yang artinya tingginya nilai F1-score yang dihasilkan membuktikan bahwa model mampu bekerja dengan lebih baik dan mampu mengurangi prediksi kelas *false*.

4. Kesimpulan

Berdasarkan pengujian model dan analisis evaluasi *confusion matrix* dapat ditarik kesimpulan bahwa algoritma *XGBoost* menghasilkan nilai akurasi lebih baik dibandingkan *AdaBoost*, *Gradient Boosting* dan *LightGBM*. Nilai akurasi yang dihasilkan adalah 92.30%. peningkatan

akurasi juga didasarkan oleh adanya pengaturan hyperparameter seperti *n_estimator*, *learning rate*, dan *max_depth*. Dari hasil ini membuktikan bahwa Teknik *boosting* mampu menangani beragamnya ulasan pelanggan. Nilai akurasi tinggi memberikan kepastian bahwa hasil klasifikasi ulasan pelanggan yang dihasilkan dapat digunakan untuk informasi baik buruknya ulasan pelanggan terhadap suatu produk *e-commerce*. Pada penelitian selanjutnya ditujukan untuk peningkatan Teknik *boosting* seperti *AdaBoost* dan Gradient Boosting. Pada algoritma *AdaBoost* masih rendahnya nilai akurasi yang dihasilkan. Pada algoritma *Gradient Boosting* nilai true negative melebihi true positive. Perlu adanya peningkatan model seperti penerapan *upsampling* untuk menyeimbangkan dataset.

Daftar Pustaka

- [1] Arif Indrawan Putra and Anita Diana, "Perancangan E-Commerce dengan Business Model Canvas untuk Peningkatan Penjualan pada Toko Parfum," *Jurnal Telematika*, vol. 15, no. 1, 2020.
- [2] H. Al Mashalah, E. Hassini, A. Gunasekaran, and D. Bhatt (Mishra), "The impact of digital transformation on supply chains through e-commerce: Literature review and a conceptual framework," *Transp Res E Logist Transp Rev*, vol. 165, p. 102837, Sep. 2022, doi: 10.1016/j.tre.2022.102837.
- [3] M. Oase Ansharullah, W. Agustin, L. Lusiana, J. Junadhi, S. Erlinda, and F. Zoromi, "Product Classification Based on Categories and Customer Interests on the Shopee Marketplace Using the Naïve Bayes Method," *JAIA-Journal Of Artificial Intelligence And Applications*, vol. 2, no. 2, pp. 15–22, 2022.
- [4] L. Donati, E. Iotti, G. Mordonini, and A. Prati, "Fashion Product Classification through Deep Learning and Computer Vision," *Applied Sciences*, vol. 9, no. 7, p. 1385, Apr. 2019, doi: 10.3390/app9071385.
- [5] R. E. Bawack, S. F. Wamba, K. D. A. Carillo, and S. Akter, "Artificial intelligence in E-Commerce: a bibliometric study and literature review," *Electronic Markets*, vol. 32, no. 1, pp. 297–338, Mar. 2022, doi: 10.1007/s12525-022-00537-z.
- [6] A. Patra, V. Vivek, B. R. Shambhavi, K. Sindhu, and S. Balaji, "Product Classification in E-Commerce Sites," in *Advanced Computing and Intelligent Engineering*, 2021, pp. 485–495. doi: 10.1007/978-981-33-4299-6_40.
- [7] L. Tan, M. Y. Li, and S. Kok, "E-Commerce Product Categorization via Machine Translation," *ACM Trans Manag Inf Syst*, vol. 11, no. 3, pp. 1–14, Sep. 2020, doi: 10.1145/3382189.
- [8] A. Noor and M. Islam, "Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2019, pp. 1–6. doi: 10.1109/ICCCNT45670.2019.8944436.
- [9] Stephenie, B. Warsito, and A. Prahatama, "Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method," *E3S Web of Conferences*, vol. 202, p. 16006, Nov. 2020, doi: 10.1051/e3sconf/2020216006.
- [10] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, "Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes," *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, vol. 6, no. 1, p. 98, Jan. 2023, doi: 10.53513/jsk.v6i1.7400.
- [11] D. L. Rianti, Y. Umaidah, and A. Voutama, "Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 1, p. 98, Aug. 2021, doi: 10.30998/string.v6i1.9993.
- [12] I. S. Milal, M. H. M. Hasanudin, M. A. Nur Azhari, R. A. Nugraha, N. Agustina, and S. E. Damayanti, "KLASIFIKASI TEKS REVIEW PADA E-COMMERCE TOKOPEDIA MENGGUNAKAN ALGORITMA SVM," *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, vol. 5, no. 1, pp. 34–45, Jun. 2023, doi: 10.53580/naratif.v5i1.191.
- [13] N. Istiqamah and M. Rijal, "Klasifikasi Ulasan Konsumen Menggunakan Random Forest dan SMOTE," *Journal of System and Computer Engineering (JSCE)*, vol. 5, no. 1, pp. 66–77, Jan. 2024, doi: 10.61628/jsce.v5i1.1061.
- [14] I. R. Hendrawan, E. Utami, and A. D. Hartanto, "Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 143–149, Jun. 2022, doi: 10.29408/edumatic.v6i1.5613.
- [15] R. V. A. Ongutu, R. M. Rimuru, and C. Otieno, "Target Sentiment Analysis Ensemble for Product Review Classification," *Journal of Information Technology Research*, vol. 15, no. 1, pp. 1–13, Aug. 2022, doi: 10.4018/JITR.299382.
- [16] S. N. Singh and T. Sarraf, "Sentiment Analysis of a Product based on User Reviews using Random Forests Algorithm," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, Jan. 2020, pp. 112–116. doi: 10.1109/Confluence47617.2020.9058128.
- [17] W. Zhao, "Classification of Customer Reviews on E-commerce Platforms Based on Naive Bayesian Algorithm and Support Vector Machine," *J Phys Conf Ser*, vol. 1678, no. 1, p. 012081, Nov. 2020, doi: 10.1088/1742-6596/1678/1/012081.
- [18] N. L. Putri, B. Warsito, and B. Surarso, "Pengaruh Klasifikasi Sentimen Pada Ulasan Produk Amazon Berbasis Rekayasa Fitur dan K-Nearest Neighbor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 65–74, Feb. 2024, doi: 10.25126/jtiik.20241117376.
- [19] M. Fayaz, A. Khan, J. U. Rahman, A. Alharbi, M. I. Uddin, and B. Alouffi, "Ensemble machine learning model for classification of spam product reviews," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8857570.
- [20] Md. M. R. Mamun, O. Sharif, and M. M. Hoque, "Classification of Textual Sentiment Using Ensemble Technique," *SN Comput Sci*, vol. 3, no. 1, p. 49, Jan. 2022, doi: 10.1007/s42979-021-00922-z.
- [21] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
- [22] J. R. Bertini Junior and M. do C. Nicoletti, "An iterative boosting-based ensemble for streaming data classification," *Information Fusion*, vol. 45, pp. 66–78, Jan. 2019, doi: 10.1016/j.inffus.2018.01.003.