

Analisis Sentimen Ulasan Pelanggan Tokopedia Menggunakan Metode *Stacked Generalization*

1st Fathih Adawi Ahmad
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
fathihadawi@student.telkomuniversity.ac.id

2nd Jondri
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
jondri@telkomuniversity.ac.id

3rd Kemas Muslim Lhaksana
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
kemasmuslim@telkomuniversity.ac.id

Abstrak-*E-commerce* merupakan salah satu bentuk perkembangan dunia khususnya perdagangan barang dan jasa. Sistem perdagangan telah berevolusi dengan hadirnya *e-commerce* yang membuat penggunaannya tidak terlepas dari teknologi internet. Produk yang tersedia di dalam *e-commerce* begitu banyak dan bervariasi macamnya. Seorang pelanggan dapat membagikan pengalamannya setelah bertransaksi, dengan menuliskan ulasan pada produk baru yang telah dibeli. Setiap ulasan menggambarkan kepuasan pelanggan dan emosi yang tertuang dalam sebuah kalimat yang belum tentu dimengerti hanya dengan membaca kata per kata saja. Faktor-faktor tersebut menjadi pembahasan dalam penelitian ini, dengan mengkategorikan analisis sentimen menjadi positif dan negatif. Pembobotan TF-IDF dalam *text classification* oleh metode *Stacked Generalization* bertujuan untuk mengetahui minat pembeli dalam membeli sebuah produk dari ulasan pembeli sebelumnya di Tokopedia. Hasil analisis pengujian menunjukkan pemodelan *stacking* mendapatkan *macro average* dengan nilai 0,67. Pengujian dilakukan pada *dataset* dengan 4.049 ulasan yang berisikan 3.551 sentimen positif dan 498 sentimen negatif.

Kata kunci- *stacked generalization, e-commerce, analisis sentimen*

Abstract-E-commerce is one form of world development, especially trade in goods and services. The trading system has evolved with the presence of e-commerce which makes its users inseparable from the name of internet technology. The products available in e-commerce are so many and varied. A customer can share his experience after a transaction, by writing a review on a new product that has been purchased. Each picture depicts customer satisfaction and emotions contained in a sentence that is not necessarily difficult just by reading word for word. These factors are discussed in this study, by categorizing the review sentiment analysis into positive and negative. The weighting of TF-IDF in text classification using the Stacked Generalization method aims to determine buyer interest in buying a product from previous buyer reviews on Tokopedia. The results of the analysis show that stacking modeling gets a macro average with a value of 0.67. The test was carried out on a dataset with 4,049

reviews containing 3,551 positive sentiments and 498 negative sentiments.

Keywords- Stacked Generalization, E-commerce, Sentiment Analysis

I. PENDAHULUAN

A. Latar Belakang

Dengan kemajuan teknologi, adanya Internet telah mewujudkan dunia maya di dunia nyata, dalam sebuah jaringan global, mempertemukan semua individu ke seluruh dunia. Salah satu contohnya dalam wujud dunia perdagangan dengan hadirnya *Electronic Commerce* atau *E-Commerce* merupakan sebuah revolusi sistem jual dan beli pada barang dan jasa. *Platform e-commerce* telah memikat banyak pelanggan karena kelebihan dan kemudahannya dalam bertransaksi, sehingga banyak pelanggan beralih dari metode belanja konvensional ke pembelian online. Terlepas dari semua kelebihan, tidak luput kekurangan yang ada dalam *e-commerce* ini. Contohnya dalam menentukan kualitas produk, adakalanya foto yang terdapat pada halaman *website* bisa berbeda dengan aslinya [1]. Selain itu, ada cara lain dalam memastikan kualitas dari sebuah produk. Pelanggan bisa membaca atau melihat ulasan dari pelanggan, khusus banyaknya pelanggan membagikan pengalamannya membeli suatu produk.

Tokopedia merupakan salah satu *e-commerce* yang ada di Indonesia. Sebelum membeli produk di Tokopedia.com, pembeli bisa melihat ulasan produk yang telah diberikan dari pembeli sebelumnya [2]. Itu merupakan salah satu dari hal yang menarik pembeli, sehingga calon pembeli dapat menilai apakah produk tersebut bagus atau tidak ketika hendak membeli. Ketika sudah membeli sebuah produk, umumnya pembeli dapat memberikan rating dan ulasan untuk memberikan feedback kepada penjual. Saat pembeli menulis ulasan, isi dari ulasan tersebut mengandung nilai positif ataupun negatif. Karena ulasan yang diberikan pembeli bersifat subjektif dimana setiap pembeli bebas untuk menuliskan ulasannya.

Berbagai penelitian sudah dilakukan sebelumnya mengenai analisis sentimen. Sebuah penelitian analisis review film dengan menggunakan algoritma *Support Vector Machine* (SVM) diikuti dengan Information Gain (IG) [3]. Pendekatan lainnya, pada penelitian dengan 3 pemodelan diantaranya *text-CNN*, *BI-LSTM*, dan *BERT* [4]. 4 kombinasi yang berbeda pada algoritma *Random Forest* [5]. Lain halnya pada penelitian dari (Irwansyah Saputra dkk, 2021) dengan menggunakan beberapa metode klasifikasi *Naïve Bayes* (NB), *Support Vector Machine* (SVM), dan *K-Nearest Neighbour* (K-NN) untuk mengetahui opini pelanggan *marketplace* di media sosial Twitter [6].

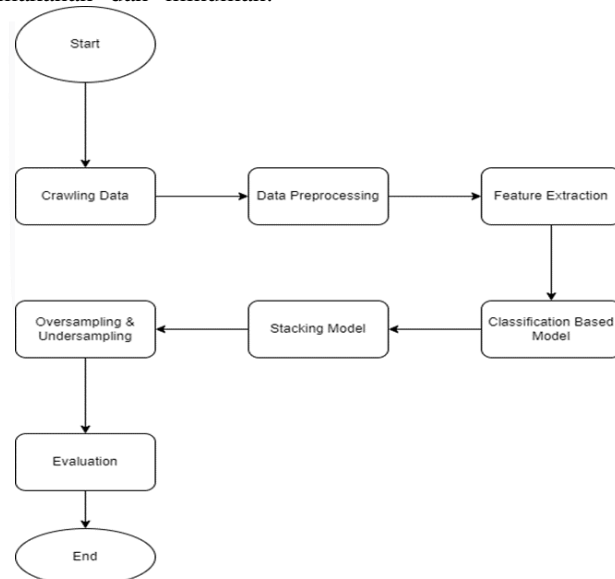
Terkait penelitian sebelumnya, berbagai algoritma *supervised learning* telah digunakan untuk mengklasifikasi. Namun, hanya menggunakan algoritma tunggal saja, tanpa adanya peleburan kombinasi pengklasifikasian. *Stacked Generalization* atau *Stacking* merupakan metode *ensemble learning* dengan menggunakan kombinasi dari beberapa algoritma. Tujuannya yaitu untuk meningkatkan kinerja analitik prediktif. Maka untuk melakukan pendekatan analisis sentimen pada serangkaian ulasan pelanggan Tokopedia terhadap produk dalam kategori makanan dan minuman.

Sesuai keinginan peneliti untuk mengklasifikasi setiap ulasan pelanggan menjadi ulasan positif dan negatif. Fokus utama bahwa penggunaan fitur TF-IDF dengan metode *Stacked Generalization* dapat menganalisis sentimen ulasan pelanggan Tokopedia, dan juga beberapa pendekatan yang akan dibahas dalam penelitian ini.

II. METODE

A. Alur Sistem

Alur sistem yang dibangun untuk analisis sentimen ulasan pelanggan Tokopedia dengan menggunakan metode *Stacked Generalization* yang terdiri dari beberapa tahapan yaitu *crawling data*, data *preprocessing*, ekstraksi fitur, pemodelan dengan model dasar, pemodelan *Stacking*, dan evaluasi. Perancangan alur sistem yang akan dilakukan dapat dilihat pada gambar 1 berikut:



GAMBAR 1
FLOWCHART ALUR SISTEM

B. Crawling Data

Pada tahap *crawling data* dimaksudkan untuk mendapatkan *dataset* ulasan, dengan menggunakan teknologi *web crawling* yaitu *web driver Selenium* yang didapatkan dari situs *web Tokopedia.com*. Dengan target URLs sebanyak 544 halaman yang mengandung detail produk dan ulasannya. Pada setiap ulasan mengandung 2 set informasi yaitu teks ulasan, dan produk rating. *Rating* berupa nilai diskrit contohnya 3, dengan rentang 0 sampai 5, untuk nilai 2.5 itu tidak bisa menjadi nilai yang berkelanjutan.

Lalu membuat skor pada rating, nilai skor 0 menandakan untuk ulasan dengan rating 0 dan begitu seterusnya sampai mendapatkan nilai skor 5 menandakan ulasan dengan rating 5.

C. Data Preprocessing

Pada tahap data *preprocessing* data mentah diubah menjadi format data yang dibutuhkan yaitu csv. Data berupa atribut yang tidak dibutuhkan akan dihapus. Untuk lebih jelasnya, terdapat 5 tugas

dalam *preprocessing* ini diantaranya sebagai berikut:

1. *Case Folding*: Mengkonversi menjadi kata-kata ke huruf kecil contohnya, "BAGUS" menjadi "bagus" dengan menggunakan library NLTK. Jadi semua huruf yang terdapat huruf besar akan diseragamkan menjadi huruf kecil.
2. *Data Cleaning*: membersihkan dari karakter yang berisikan emoticon, angka, tanda baca yang tidak digunakan. Tujuan dari proses ini untuk mengurangi *noisy data*.
3. Tokenisasi: Teknik untuk memisahkan kalimat menjadi kata-kata individu. Seperti simbol, kata kunci, frasa yang dikenal sebagai token. Dan semua simbol karakter seperti tanda seru, titik koma akan dihilangkan.
4. *Cleaning stop words*: *Stop words* adalah istilah yang tidak diperlukan dalam sebuah kalimat. Pada setiap ulasan yang mengandung stop word, istilah yang masuk ke dalam *stop words* pada ulasan tersebut akan dihapus untuk meningkatkan efisiensi analisis. Untuk implementasinya akan menggunakan *corpus library*.
5. *Stemming*: Proses *stemming* digunakan untuk menormalisasikan suatu kata menjadi bentuk kata dasarnya. Untuk mendukung proses ini juga diterapkan *library* sastrawi.

Setelah tahap *preprocessing*, selanjutnya proses *labelling* secara manual. Data yang sudah terseleksi sebelumnya masuk ke dalam proses *labelling* sentimen. Proses *labelling* sentimen ini, dilakukan secara manual dimana keseluruhan kalimat akan dibaca dan dipahami apakah kalimat tersebut masuk ke dalam sentimen negatif atau positif. Untuk mengurangi hasil yang subjektif maka

proses ini dilakukan oleh 3 orang, dengan metode penilaian *voting*.

D. Ekstraksi Fitur

Tahap ekstraksi fitur harus menggambarkan data dalam format yang diperlukan oleh algoritma pembelajaran mesin, sehingga bisa digunakan untuk menyelesaikan untuk tugas. Sesuai dalam penelitian ini dapat mengidentifikasi emosi dari ulasan pelanggan. Ini semua dilakukan dengan menggunakan algoritma TF-IDF. Ekstraksi dilakukan menggunakan vektorisasi TF-IDF. Untuk formulanya berikut di bawah ini [7].

$$TFIDF_{ij} = TF_{ij} \times idf_i \tag{1}$$

$$TF_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \tag{2}$$

$$IDF_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \tag{3}$$

n_{ij} merupakan jumlah kata i dalam dokumen j . $\sum_k n_{kj}$ total kata dalam dokumen yang muncul. $|D|$ total kata dalam kumpulan tulisan. Dan $|\{j: t_i \in d_j\}|$ ialah jumlah kumpulan tulisan yang termasuk dalam t_i .

Hasilnya berisi set data latih yang didasarkan pada parameter fitur maksimum, di mana mewakili jumlah maksimum kata unik untuk dikenali dalam keseluruhan ulasan. Prosesnya itu sebuah kalimat berisikan kumpulan kata dipisah menjadi kata per kata dan mengubahnya menjadi bilangan unik. Contohnya misalkan dokumen 1 "Produk sesuai ekspektasi dan pengiriman kilat" lalu akan di tokenisasi menjadi ['produk', 'sesuai', 'ekspektasi', 'dan', 'pengiriman', 'kilat'] kemudian dikonversi sesuai dengan hasil pembobotannya dari setiap kata [1, 2, 3, 4, 5, 6]. Jadi 2 menggambarkan dari kata sesuai, 6 menggambarkan kata kilat. Kemudian *dataset* dipisah menjadi data primer dan eksperimen untuk dilatih oleh pemodelan. Berikut ilustrasi dari penerapan TF-IDF pada tabel 1 di bawah ini.

TABEL 1
ILUSTRASI TF-IDF

| Dokumen | Isitilah 0 | Isitilah 1 | Isitilah 2 | ... | Istilah (n) |
|-------------|------------|------------|------------|-----|-------------|
| 0 | 5 | 6 | 0 | ... | 0 |
| 1 | 0 | 0 | 2 | ... | 0 |
| ... | ... | ... | ... | ... | ... |
| Dokumen (n) | 0 | 1 | 2 | ... | 3 |

D. Klasifikasi Model Dasar

Setelah ekstraksi fitur, pada tahap ini kemudian diterapkan algoritma pengklasifikasian diantaranya *Support Vector Machine*, *Naïve Bayes*, *Decision Tree*. Nantinya Algoritma yang sudah disebutkan sebelumnya dapat merepresentasikan probabilitas

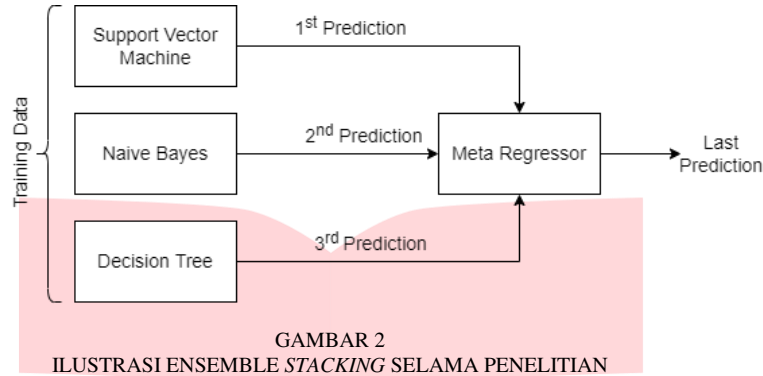
fitur dengan menggunakan data latih 80% dan 20% data uji. Algoritma *Naïve Bayes* dipilih karena mudah diimplementasikan dan cepat [8]. *Support Vector Machine* salah satu pengklasifikasian yang sederhana dan dapat memberikan akurasi tinggi [9]. *Decision Tree* merupakan salah satu teknik data

mining yang populer [10]. Untuk implementasi *Support Vector Machine*, *Naïve Bayes*, *Decision Tree*, library *scikit-learn* pada *python* digunakan untuk memudahkan selama penelitian.

E. *Stacking Model*

Pada tahap ini, setelah pemodelan dengan algoritma *based model* selesai maka akan dilanjutkan dengan pemodelan *Stacking*. Pada tahap

pertama, menggunakan *Support Vector Machine*, *Naïve Bayes*, *Decision Tree* sebagai regresi tingkat pertama. Setelah itu, dengan algoritma *combiner* untuk membuat prediksi akhir berdasarkan semua prediksi yang digunakan tahap pertama. Algoritma *Logistic Regression* digunakan untuk mendapatkan prediksi yang lebih baik dari *based learner*. Berikut skema *Stacked Generalization* yang diimplementasikan dalam penelitian ini.



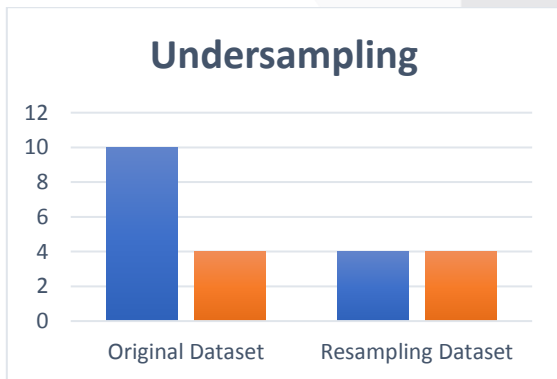
F. *Oversampling* dan *Undersampling*

Suatu algoritma pembelajaran mesin menghasilkan kinerja yang baik ketika jumlah data pada setiap kelas sama atau seimbang. Namun lain halnya pada kasus imbalanced data. *Imbalanced* data merupakan hal yang sering ditemui ketika jumlah perbandingan disuatu kelas lebih tinggi dari kelas lainnya. Ini menjadi *problematic* terhadap hasil performansi pengklasifikasian. Salah satu solusi yang biasa diterapkan yaitu, dengan menaikkan jumlah data pada kelas yang lebih rendah dan menghapus sejumlah kelas yang lebih tinggi untuk diseimbangkan dengan kelas lainnya. Berikut pendekatan untuk mengatasi permasalahan ketidakseimbangan pada variansi kelas di bawah ini.

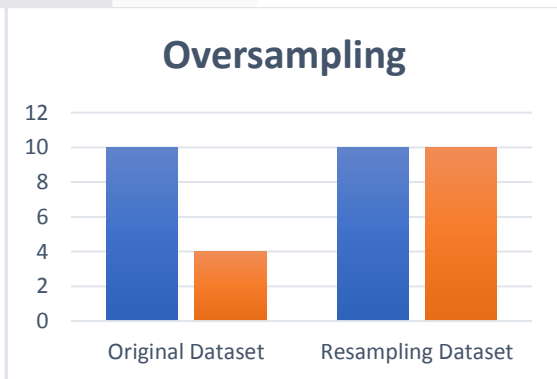
1. *Oversampling* merupakan teknik menduplikasi sample data pada kelas minoritas.
2. *Undersampling* merupakan teknik menghapus sample data pada kelas mayoritas.

Bagaimanapun algoritma pembelajaran mesin tetap menghasilkan akurasi yang tinggi namun pada kelas mayoritasnya saja, tetapi tidak memaksimalkan pada kelas minoritas. Untuk itu pada penelitian ini dilakukan pendekatan *Oversampling* dan *Undersampling* dalam mengatasi permasalahan yang ditemui yaitu *imbalanced* data. Sehingga hasil yang didapatkan dari tiap pengklasifikasian menjadi lebih baik. Berikut adalah ilustrasi dari

Oversampling dan *Undersampling* dapat dilihat pada gambar 3 dan 4.



GAMBAR 3. ILUSTRASI PENDEKATAN *UNDERSAMPLING*.



GAMBAR 4. ILUSTRASI PENDEKATAN *OVERSAMPLING*.

G. Evaluasi

Tahap evaluasi untuk mengukur performansi pengklasifikasian. Juga penggunaan evaluasi untuk membandingkan performansi *base learner* dengan *meta-learner*. Keakuratan pengklasifikasian pada *dataset* yang diberikan adalah persentase yang diklasifikasi dengan benar oleh pengklasifikasian [11]. Tiga ukuran yang biasa digunakan *Precision*, *Recall*, *F1-score*. Sebelum membahas pengukurannya, ada beberapa istilah yang perlu diketahui sebagai berikut:

1. *True Positive* (TP): menyatakan jumlah data yang positif diklasifikasi dengan benar.
2. *False Positive* (FP): menyatakan jumlah data yang positif diklasifikasi tidak akurat.
3. *False Negative* (FN): menyatakan jumlah data yang negatif diklasifikasi tidak akurat.
4. *True Negative* (TN): menyatakan jumlah data yang negatif diklasifikasi dengan benar.

Precision: Hasil kalkulasi *precision* mengembalikan berapa banyak jumlah catatan yang diklasifikasi dengan akurat. Persamaannya ditulis sebagai berikut:

$$P = TP / (TP + FP) \tag{4}$$

Recall: Adalah rasio jumlah instance yang diklasifikasi secara akurat dengan jumlah instance yang diprediksi. Semakin tinggi nilai hasil *recall* menunjukkan lebih sedikit data negatif palsu. Persamaannya ditulis sebagai berikut:

$$R = TP / (TP + FN) \tag{5}$$

F1-Score: Menggabungkan presisi dan *recall* lalu menghasilkan metrik tunggal yaitu *F1-score*. Persamaannya ditulis sebagai berikut:

$$F = (2 \times P \times R) / (P + R) \tag{6}$$

Accuracy: kalkulasi seberapa sering pengklasifikasian membuat prediksi dengan benar. Persamaannya ditulis sebagai berikut:

$$Accuracy = (Correct Prediction) / (Total data point) \tag{7}$$

III. HASIL DAN PEMBAHASAN

A. Hasil Riset

Pada proses *crawling* data menghasilkan data yang tidak terstruktur sehingga dapat menghambat proses dari pemodelan. Maka dari itu tahap *preprocessing* ini membantu untuk mendapatkan struktur data yang baik dengan menghapus segala bentuk karakter berupa *emoticon*, tanda baca, *stop words*, dan mengubah penggunaan huruf kapital tahap *preprocessing* terdapat data *null* atau data kosong, duplikasi data, dan ulasan yang berisikan *emoticon* saja, sehingga tidak terdapat kalimat dari ulasan pelanggan tersebut.

Dari semua hal tersebut mengandung informasi yang tidak berharga atau *noise data* yang dapat mengganggu selama pengklasifikasian berlangsung. Dari total keseluruhan data ulasan sebanyak 5.428, setelah dilakukan *preprocessing* jumlah data menyusut menjadi 4.049, dengan jumlah sentimen positif sebanyak 3.551 ulasan dan sentimen negatif sebanyak 498 ulasan. Hasil *preprocessing* ditunjukkan pada tabel 2 dan jumlah sentimen dapat dilihat pada gambar 5 di bawah ini.

TABEL 2.
SEBELUM DAN SESUDAH *PREPROCESSING* SENTIMEN

| Sebelum | Sesudah |
|--|---|
| Enaak dan guriih kacangnya 🍪 | ['enaak', 'guriih', 'kacang'] |
| keripik nya sudah sampe semua komplit enak dan langganan sejati 🍪 sukses keripik riki | ['keripik', 'nya', 'lengkap', 'enak', 'langgan', 'sejati', 'sukses', 'keripik', 'riki'] |
| pilus cikir alot dan liat pengiriman lama | ['pilus', 'cikir', 'alot', 'lihat', 'kirim'] |
| Pedasnya pas.cuma gak berasa Singkong nya" saran aja, kalau pembelian 5 bungkus" ada bonus.. biar makin lariss 😊 | ['pedas', 'pas', 'asa', 'singkong', 'nya', 'saran', 'beli', 'bungkus', 'bonus', 'biar', 'lariss'] |



GAMBAR 5.
TOTAL JUMLAH DATA ULASAN BERLABEL

Selanjutnya dari *preprocessing* masuk ke dalam proses ekstraksi fitur menggunakan TF-IDF. Sebelum masuk ke dalam fitur ekstraksi data dilakukan *splitting* untuk mendapatkan data latih dan uji. Pada penelitian ini menggunakan perbandingan 80% data latih dan 20% data uji. Kemudian masuk ke dalam proses TF-IDF, setiap kata diekstraksi didapatkan pembobotan kata yang sesuai. Pada penelitian menggunakan (*min_df*) berguna untuk menghapus istilah yang jarang muncul dengan presentasi sebesar 0.01 dimana istilah yang kurang dari 1% keseluruhan dokumen dan (*max_df*) berguna untuk menghapus istilah yang sering muncul dengan presentase sebesar 0.90 atau 90% istilah dari keseluruhan dokumen.

B. Hasil dan Evaluasi Klasifikasi

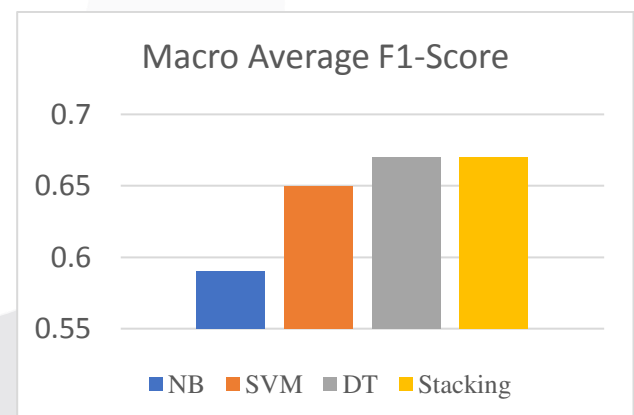
Hasil dari TF-IDF kemudian masuk ke dalam tahapan pemodelan oleh masing-masing algoritma *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, dan *Stacking* model. Setiap ulasan positif dan negatif diklasifikasi oleh tiap pemodelan untuk menguji hasil dari data yang sudah diperoleh, apakah dapat mengklasifikasi dengan baik atau tidak. Berikut adalah perbandingan hasil evaluasi macro average setiap pemodelan dasar yang ditunjukkan pada tabel 3.

TABEL 3
HASIL EVALUASI 1

| Method | Macro Average | | |
|----------|---------------|--------|------------|
| | Precision | Recall | F1-Measure |
| NB | 0,86 | 0,57 | 0,59 |
| SVM | 0,85 | 0,61 | 0,65 |
| DT | 0,7 | 0,65 | 0,67 |
| Stacking | 0,82 | 0,63 | 0,67 |

C. Analisis

Dari hasil evaluasi pada gambar 6, hasil *macro average F1-Score* terbaik terletak pada pemodelan *Stacking model* dan *Decision Tree* dengan nilai masing-masing 0,67. Dan nilai terkecilnya terletak pada pemodelan *Naïve Bayes* dengan 0,59. Pada tabel 3 pembobotan pada nilai *macro average precision* dan *recall* untuk masing-masing pemodelan perbandingannya tidak begitu bagus dikarenakan jumlah sentimen positif lebih banyak ketimbang dengan jumlah sentimen negatifnya.

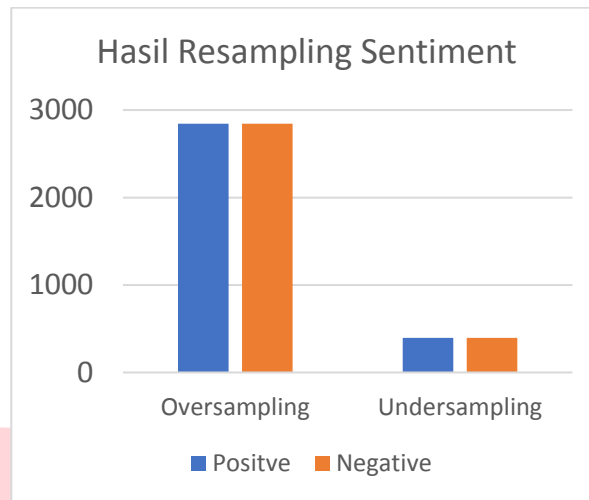


GAMBAR 6.
HASIL MACRO AVERAGE F1-SCORE KLASIFIKASI

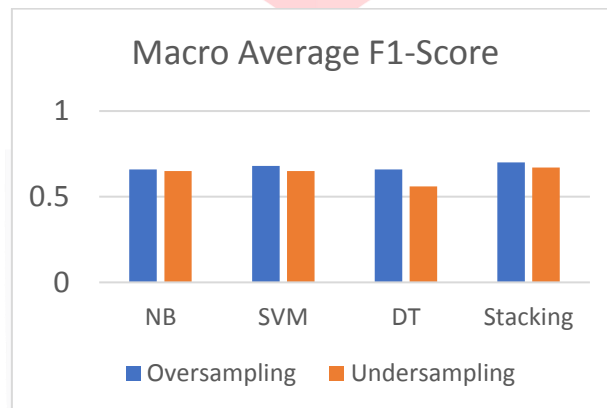
IV. PENANGANAN IMBALANCED DATA

Sebelumnya besaran data ulasan sentimen negatif terlihat terpaut jauh jumlahnya dengan data ulasan sentimen positif. Bias pada kumpulan data ini dapat mempengaruhi performansi dari algoritma pembelajaran mesin, sehingga data mayoritas disini lebih diutamakan ketimbang data minoritasnya. Untuk itu dilakukan pendekatan *Oversampling* dan *Undersampling* untuk mengatasi imbalance pada *dataset*. Pada gambar 7 menunjukkan jumlah data *Oversampling* sentimen label positif dan negatif masing-masing sebesar 2.843 dan data *Undersampling* sentimen label positif dan negatif

masing-masing sebesar 396, keduanya dilakukan pada data latih.



GAMBAR 7. HASIL OVERSAMPLING DAN UNDERSAMPLING DATASET



GAMBAR 8. HASIL MACRO AVERAGE F1-SCORE KLASIFIKASI 2

Dari hasil yang ditunjukkan pada gambar 8 dengan menerapkan pendekatan Oversampling dan Undersampling dari hasil pemodelan yang meraih nilai macro average *F1-Score* terbaik yaitu *Stacking model* dengan 0,67 dan 0,7. Sedangkan model yang mendapat nilai terkecil macro average *F1-Score* pada pendekatan *Oversampling* ialah model *Naïve*

Bayes dan *Decision Tree* dengan 0,66 dan nilai terkecil pada pendekatan *Undersampling* ialah model *Decision Tree* dengan 0,56. Untuk lebih detailnya berikut hasil evaluasi pendekatan *Oversampling* dan *Undersampling* dari tiap model yang ditunjukkan pada tabel 4 dan 5.

TABEL 4. HASIL EVALUASI MACRO AVERAGE 2

| Approach | Method | Macro Average | | |
|--------------|----------|---------------|--------|------------|
| | | Precision | Recall | F1-Measure |
| Oversampling | NB | 0,64 | 0,76 | 0,66 |
| | SVM | 0,66 | 0,78 | 0,68 |
| | DT | 0,64 | 0,71 | 0,66 |
| | Stacking | 0,67 | 0,75 | 0,7 |

TABEL 5.
HASIL EVALUASI *MACRO AVERAGE* 3

| Approach | Method | Macro Average | | |
|---------------|----------|---------------|--------|------------|
| | | Precision | Recall | F1-Measure |
| Undersampling | NB | 0,64 | 0,76 | 0,65 |
| | SVM | 0,64 | 0,79 | 0,65 |
| | DT | 0,58 | 0,68 | 0,56 |
| | Stacking | 0,66 | 0,8 | 0,67 |

Pada tabel 4 dan 5 nilai *macro average* pada keseluruhan untuk *Oversampling* lebih baik dari *Undersampling* yang cenderung sedikit lebih kecil pada masing-masing pemodelan. Ini menunjukkan dari hasil pendekatan *Oversampling* itu sendiri yaitu meningkatkan variansi pada *dataset* dengan penambahan duplikasi pada kelas mayoritasnya, sedangkan *Undersampling* sedikit lebih kecil dikarenakan variansi pada *dataset* dengan mengurangi kelas mayoritasnya.

V. KESIMPULAN

Dalam penelitian ini, penulis bertujuan untuk mengklasifikasi ulasan pelanggan *e-commerce* Tokopedia menjadi sentimen positif dan negatif, khususnya pada produk kategori makanan dan minuman. Ulasan sentimen positif dan negatif diklasifikasi oleh tiga algoritma *supervised learning* dibandingkan hasilnya dengan metode *ensemble learning*. Metode *ensemble learning* menggunakan *Stacking model* dikombinasikan dengan *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, dan *final estimator* yaitu *Logistic Regression*. Semua eksperimen dalam penelitian ini dilakukan dengan skenario yang berbeda-beda. Diantaranya klasifikasi oleh masing-masing *supervised learning* dan *ensemble learning*, kemudian pendekatan *Oversampling* dan *Undersampling* untuk mengatasi *imbalance* data. Hasil penelitian menunjukkan bahwa teknik *ensemble learning* dengan *Stacking model* mendapatkan hasil yang lebih baik dibandingkan pengklasifikasian oleh metode *supervised learning*. Dengan nilai *macro average F1-Score* 0,67 lalu pendekatan *Oversampling* mendapatkan 0,70 dan terakhir pendekatan *Undersampling* sebesar 0,67.

REFERENSI

- [1] A. Alrehili and K. Albalawi, "Sentiment Analysis of Customer Reviews Using Ensemble Method," International Conference on Computer and Information Sciences (ICIS), pp. 1-6, 2019.
- [2] F. V. Sudjatmika, "Pengaruh Harga, Ulasan Produk, Kemudahan, dan Keamanan Terhadap Keputusan Pembelian Secara Online di Tokopedia. Com," AGORA Vol. 5, No. 1, 2017.
- [3] R. Maulana, P. A. Rahayuningsih, W. Irmayani, D. Saputra, and W. E. Jayanti, "Improved accuracy of sentiment analysis movie review using Support Vector Machine based Information Gain," J. Phys. Conf. Ser., vol. 1641, no. 1, p. 012060, 2020.
- [4] H. Zikang, Y. Yong, Y. Guofeng and Z. Xinyu, "Sentiment analysis of agricultural product ecommerce review data based on deep learning," 2020 International Conference on Internet of Things and Intelligent Applications (ITIA), pp. 1-7, 2020.
- [5] A. Ejaz, Z. Turabee, M. Rahim and S. Khoja, "Opinion mining approaches on Amazon product reviews: A comparative study," pp. 173-179, 2017.
- [6] I. Saputra et al., "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning," Fakt. exacta, vol. 13, no. 4, pp. 200-207, 2021.
- [7] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," Expert Systems with Applications, vol. 36, no. 3, pp. 6527–6535, 2009.
- [8] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1275-1279, 2018.
- [9] J. Jabbar, I. Urooj, W. JunSheng, and N. Azeem, "Real-time sentiment analysis on E-commerce application," IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), pp. 391-396, 2019.
- [10] Zhang Yang, Liu Aodong, Li Maoqing, "Decision Tree Algorithm Based on the Information Theory", Control Theory and

Applications, Vol. 25, Issue 1, pp. 4-7, 2006.

- [11] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," IEEE International Conference on Innovative Research and Development (ICIRD), pp. 1-6, 2018.

