

Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia

(Aspect Category Classification with Machine Learning Approach Using Indonesian Language Dataset)

Syaifulloh Amien Pandega Perdana¹, Teguh Bharata Aji², Ridi Ferdiana³

Abstract—Customer reviews are opinions on the quality of goods or services that consumers perceive. Customer reviews contain useful information for both consumers and providers of goods or services. The availability of a large number of customer reviews on the website requires a framework for extracting sentiment automatically. A customer review often contains many aspects, so the Aspect Based Sentiment Analysis (ABSA) should be used to determine the polarity of each aspect. One of the important tasks in ABSA is Aspect Category Detection. The application of Machine Learning Methods for Aspect Category Detection has been mostly done in the English language domain, but in the Indonesian language domain, there are still a few. This study compares the performance of three machine learning algorithms, namely Naïve Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF), on Indonesian language customer reviews using Term Frequency-Inverse Document Frequency (TF-IDF) as term weighting. The results show that RF performs the best, compared to NB and SVM, in three different domains, namely restaurants, hotels, and e-commerce, with the f1-scores for each domain are 84.3%, 85.7%, and 89.3%.

Intisari—Ulasan pelanggan merupakan opini terhadap kualitas barang atau jasa yang dirasakan konsumen. Ulasan pelanggan mengandung informasi yang berguna bagi konsumen maupun penyedia barang atau jasa. Ketersediaan ulasan pelanggan dalam jumlah besar pada *website* membutuhkan suatu *framework* untuk mengekstraksi sentimen secara otomatis. Sebuah ulasan pelanggan sering kali mengandung banyak aspek sehingga *Aspect Based Sentiment Analysis* (ABSA) harus digunakan untuk mengetahui polaritas masing-masing aspek. Salah satu tugas penting dalam ABSA adalah *Aspect Category Detection*. Metode *machine learning* untuk *Aspect Category Detection* sudah banyak dilakukan pada domain berbahasa Inggris, tetapi pada domain bahasa Indonesia masih sedikit. Makalah ini membandingkan kinerja tiga algoritme *machine learning*, yaitu *Naïve Bayes* (NB), *Support Vector Machine* (SVM), dan *Random Forest* (RF) pada ulasan pelanggan berbahasa Indonesia menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai *term weighting*. Hasil menunjukkan bahwa RF memiliki kinerja paling unggul dibandingkan NB dan SVM pada tiga domain yang berbeda, yaitu restoran, hotel, dan *e-commerce*, dengan nilai *f1-score* untuk masing-masing domain adalah 84,3%, 85,7%, dan 89,3%.

Kata Kunci—*Aspect Category Classification, Naïve Bayes, Support Vector Machine, Random Forest, Machine Learning, Ulasan Pelanggan.*

^{1,2,3} Magister Teknologi Informasi, Departemen Teknik Elektro dan Teknologi Informasi, Fakultas Teknik UGM, Jl. Grafika No. 2, Kampus UGM, Yogyakarta, Indonesia 55281 (telp: 0274-552305; e-mail: ¹syaifullohamien@mail.ugm.ac.id, ²adji@ugm.ac.id, ³ridi@ugm.ac.id)

I. PENDAHULUAN

Ulasan pelanggan pada media *online* merupakan opini dari pengalaman yang didapatkan oleh konsumen terhadap jasa layanan ataupun produk. Pengalaman tersebut kemudian disampaikan melalui media umpan balik yang disediakan atau pada media *online* yang lain. Ulasan tersebut berguna bagi konsumen dan penjual produk atau penyedia jasa. Bagi konsumen, ulasan pelanggan menjadi bahan pertimbangan dalam menentukan kualitas produk yang akan dibeli atau jasa yang ingin didapatkan, sedangkan bagi penjual produk ataupun penyedia jasa, ulasan pelanggan dapat menjadi masukan untuk produk ataupun layanan jasanya sehingga dapat ditingkatkan kualitasnya [1]. Ulasan pelanggan tersebut terbukti informatif dan membantu, tetapi ulasan yang disampaikan masih mengandung informasi yang tidak lengkap, kata-katanya tidak formal, dan bersifat bias sehingga diperlukan penambangan opini atau analisis sentimen terhadap ulasan pelanggan tersebut.

Analisis sentimen merupakan salah satu topik pada *Natural Language Processing* (NLP) yang bertujuan mengekstraksi dan mengidentifikasi suatu kalimat sehingga memperoleh polaritas sentimen pada kalimat tersebut. Polaritas sentimen dari suatu kalimat dapat bernilai positif, negatif, atau netral [2]. Analisis sentimen biasanya terbatas pada satu aspek saja yang dapat ditentukan polaritas sentimennya, sedangkan dalam suatu kalimat bisa terdapat beberapa aspek yang masing-masing aspek tersebut mempunyai polaritas sentimen. Oleh karena itu, analisis yang lengkap harus dilakukan untuk dapat menentukan aspek yang terdapat pada suatu kalimat kemudian mengidentifikasi polaritas sentimen yang terdapat dalam aspek tersebut sehingga diketahui bernilai positif, negatif, atau netral. Dalam hal ini, sentimen analisis berbasis aspek atau *Aspect-Based Sentiment Analysis* (ABSA) akan diperlukan untuk dapat menentukan aspek serta polaritas sentimennya [3].

ABSA merupakan model yang mengekstraksi aspek dan sentimen dalam suatu kalimat sehingga dapat diketahui jumlah aspek dalam kalimat tersebut dan ditentukan polaritas sentimen dari masing-masing aspek. Sebagai contoh, dalam kalimat ulasan restoran, “Makanan dan tempatnya enak, akan tetapi pelayannya kurang memuaskan,” terdapat beberapa aspek, antara lain tempat, makanan, dan layanan, yang masing-masing aspek memiliki polaritas sentimen, yaitu positif untuk aspek makanan, positif untuk aspek tempat, dan negatif untuk aspek layanan.

International workshop tentang ABSA sudah dimulai dari *Semantic Evaluation* (SemEval) 2014 [4], 2015 [5], dan 2016

[6], yang memiliki tiga tugas, yaitu *Aspect Category Detection*, *Opinion Target Expression*, dan *Sentiment Polarity Classification*. *Aspect Category Detection* berusaha mengatasi masalah dalam mengelompokkan ulasan pelanggan ke dalam beberapa aspek yang aspeknya sudah ditentukan sebelumnya [7].

Banyak penelitian yang sudah mendiskusikan tentang *Aspect Category Detection* dengan menggunakan berbagai *dataset* dan pendekatan algoritme. Salah satu algoritme yang digunakan adalah *machine learning*. *Naive Bayes* (NB), *Support Vector Machine* (SVM), *Maximum Entropy* (ME), *Random Forest* (RF), dan *Fuzzy Lattice Reasoning* (FLR) merupakan beberapa metode yang terdapat pada *machine learning*. Dengan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) sebagai *term weighting* dan *dataset* ulasan hotel berbahasa Inggris, dihasilkan metode SVM sebagai yang terbaik. Sementara itu, dengan menggunakan *dataset* ulasan restoran dan *term weighting* yang sama, didapatkan RF sebagai yang terbaik [8]. Ulasan restoran berbahasa Inggris juga digunakan sebagai *dataset* penelitian lain, yang menghasilkan metode NB sebagai yang terbaik [9]. Penggunaan algoritme NB, RF, dan SVM juga pernah dilakukan dengan menggunakan dua domain yang berbeda dan menghasilkan RF sebagai algoritme yang terbaik [10].

Permasalahan yang muncul dari beberapa penelitian di atas adalah belum terdapatnya penelitian yang membandingkan beberapa algoritme pada *machine learning* untuk klasifikasi aspek dengan berbagai domain *dataset* berbahasa Indonesia. Oleh karena itu, makalah ini bertujuan mencari beberapa algoritme *machine learning* untuk klasifikasi aspek dengan *dataset* ulasan pelanggan berbahasa Indonesia dari beberapa domain. Dengan dilakukannya penelitian ini, diharapkan dapat ditemukan algoritme *machine learning* yang terbaik untuk klasifikasi aspek.

II. ASPECT CATEGORY DETECTION

Salah satu tugas dalam ABSA adalah *Aspect Category Detection*, yang bertujuan mengelompokkan ulasan pelanggan ke dalam beberapa aspek yang aspeknya sudah ditentukan sebelumnya. Jumlah aspek yang terdapat pada ulasan berbeda-beda, tergantung pada domainnya. Dalam domain ulasan restoran [11]–[13], terdapat beberapa aspek, antara lain makanan, pelayanan, harga, dan tempat. Terdapat penelitian lain dengan *dataset* bahasa Indonesia, tetapi memiliki domain yang berbeda. Pada domain belanja *online* [14], [15], terdapat tujuh aspek, yaitu akurasi, kualitas, komunikasi, waktu pengiriman, harga, pengemasan, dan komunikasi. Sementara itu, pada domain hotel [16], terdapat beberapa aspek, antara lain lokasi, kamar dan fasilitas.

Terdapat beberapa metode *term weighting*, antara lain *Term Frequency* (TF), *Inverse Document Frequency* (IDF), dan TF-IDF [17]. TF-IDF merupakan model yang sering digunakan untuk *term weighting* karena efisien, mudah, dan memiliki hasil yang akurat [18]. Setelah didapatkan matriks TF-IDF, dilakukan klasifikasi dengan algoritme *machine learning*.

Beberapa metode *machine learning* yang biasa digunakan untuk klasifikasi adalah NB, SVM, ME, RF, dan LR [8].

Dengan membandingkan kelima metode tersebut sebagai klasifikasi dan TF-IDF sebagai ekstraksi fitur, maka didapatkan hasil yang berbeda untuk dua domain *dataset*. Metode SVM menghasilkan akurasi terbaik untuk domain hotel, sedangkan untuk domain restoran, metode RF menghasilkan nilai akurasi yang terbaik.

Metode SVM juga menunjukkan kinerja paling bagus pada penelitian *Aspect Category Detection* [19], dibandingkan dengan *Decision Tree* (DT), NB, dan *k-Nearest Neighbor*. TF-IDF juga digunakan sebagai *term weighting*, kemudian klasifikasi dilakukan dengan SVM sebagai metode yang terbaik dan menghasilkan nilai akurasi 73,48%.

Penelitian lain mendapatkan hasil yang berbeda [10], dengan membandingkan enam algoritme *machine learning*, yaitu NB, DT, SVM, RF, *Extra Trees Classifier*, dan *Extreme Gradient Boosting* (XGBoost), untuk dua domain *dataset* dalam bahasa Inggris. XGBoost menunjukkan nilai tertinggi dalam hal akurasi dengan nilai 63,33% untuk *dataset* makanan dan 65,11% untuk *dataset* teknologi. Di sisi lain, penelitian dengan *dataset* berbahasa Indonesia sudah pernah dilakukan dengan domain retorasi sebagai *dataset* [12]. Dengan menggunakan *MaxEnt Classifier* untuk klasifikasi dan *Continuous Bag-of-Word* (CBOW) sebagai *feature extraction*, didapatkan nilai *f1-measure* 82,30%, *precision* 77,75%, dan *recall* 88,84%.

Berdasarkan penelitian-penelitian yang sudah pernah dilakukan, diusulkan penggunaan tiga metode algoritme *machine learning*, yaitu NB, SVM, dan RF untuk mengklasifikasikan aspek yang terdapat pada ulasan. Hal ini dilakukan karena penggunaan ketiga metode tersebut, dengan tiga domain *dataset* bahasa Indonesia, belum pernah dilakukan.

III. METODOLOGI

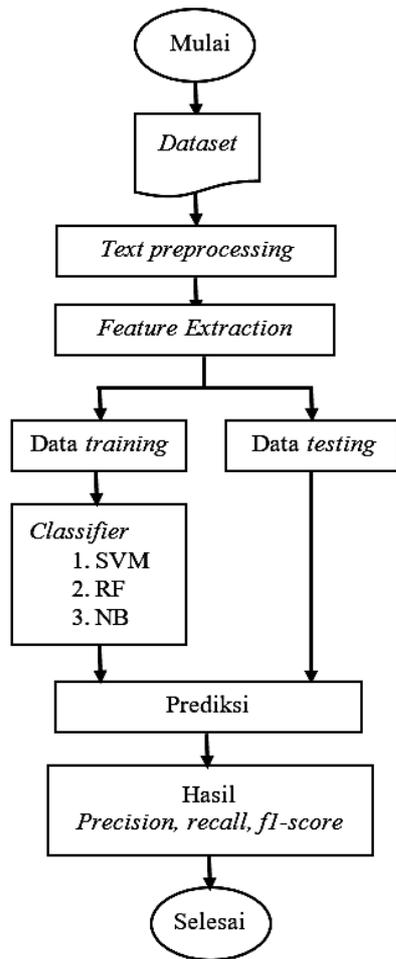
Pada makalah ini, digunakan beberapa domain *dataset* bahasa Indonesia untuk *Aspect Category Classification* dengan pendekatan algoritme *machine learning* (NB, SVM, dan RF). Gbr. 1 merupakan tahapan-tahapan yang dilakukan. Tahapan 1 mengumpulkan tiga *dataset* dengan domain restoran, *e-commerce*, dan hotel. Tahapan 2 memproses data menggunakan *text preprocessing*. Pada tahapan 3, *term weighting* mengubah teks menjadi angka dan tahapan 4 adalah klasifikasi aspek dengan menggunakan algoritme *machine learning*. Pada tahap terakhir, dihitung hasil dari klasifikasi *machine learning* dengan *precision*, *recall*, dan *f1-score*.

A. Dataset

Pada makalah ini digunakan tiga domain *dataset* berbahasa Indonesia, yaitu domain restoran, *e-commerce*, dan hotel. Pada domain restoran, sebanyak 1.521 kalimat, sesuai dengan penelitian sebelumnya [11], dibagi menjadi 1.099 kalimat data *training* dan 422 kalimat data *testing*. Domain *e-commerce* menggunakan sebanyak 572 kalimat [20], sedangkan domain hotel sebanyak 2.880 kalimat [21].

B. Text Preprocessing

Tahapan *text preprocessing* merupakan tahapan awal terhadap teks untuk mempersiapkan teks menjadi data yang siap diproses lebih lanjut. Pada tahapan ini terdapat tiga proses, yaitu *case folding*, *stopword removal*, dan *stemming*. Proses



Gbr. 1 Tahapan penelitian.

case folding mengonversi teks menjadi huruf kecil sehingga data memiliki format yang sama serta menghapus tanda baca yang tidak diperhitungkan. Selanjutnya, proses *stopword removal* menghapus kata-kata umum yang sering muncul dalam ulasan, tetapi tidak berkontribusi pada keakuratan hasil klasifikasi. Proses terakhir adalah *stemming*, yang bertujuan menghilangkan imbuhan, sisipan, dan akhiran yang terdapat pada data sehingga akan kembali ke kata dasar.

C. Term Weighting

Setelah *text preprocessing*, tahap selanjutnya adalah *term weighting*, yaitu mengonversi data teks menjadi angka. Pada makalah ini, digunakan TF-IDF sebagai *term weighting*. TF-IDF digunakan untuk memberikan nilai bobot pada setiap kata yang terdapat dalam *dataset* dengan menghitung jumlah kemunculannya. Metode ini menghitung nilai TF dan IDF pada setiap kata di setiap dokumen. Metode ini menghitung bobot setiap kata *t* di dokumen *d* sesuai dengan (1).

$$W_{dt} = t f_{dt} \times I D F_t \tag{1}$$

dengan

- d* = dokumen ke-*d*
- t* = kata ke-*t* dari kata kunci
- W* = bobot dokumen ke-*d* terhadap kata ke-*t*

TABEL I
CONFUSION MATRIX

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

T_f = banyaknya kata yang dicari pada sebuah dokumen
IDF = Inverse Document Frequency.

D. Klasifikasi

Salah satu algoritme yang digunakan untuk klasifikasi pada *Aspect Category Classification* ialah *machine learning*. NB, SVM, dan RF merupakan beberapa metode yang terdapat pada algoritme *machine learning*.

Classifier NB didasarkan pada teorema Bayes yang memprediksi probabilitas keanggotaan kelas suatu data berdasarkan perhitungan probabilitas. Metode ini digunakan karena menjanjikan untuk menyelesaikan masalah analisis data tekstual dan memiliki akurasi yang tinggi dengan perhitungan sederhana [22].

SVM merupakan metode *supervised learning* yang mengidentifikasi pola dan digunakan untuk analisis regresi serta klasifikasi. SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dalam pemodelan klasifikasi karena membuat jalur *hiper* yang paling cocok untuk memisahkan dua kelas dengan celah terbesar. SVM menentukan pemisah linear di ruang pencarian yang terbaik dalam memisahkan kelas-kelasnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun nonlinear [23].

Classifier RF merupakan salah satu metode yang digunakan dalam klasifikasi yang berdasarkan kumpulan metode pembelajaran dengan menggunakan DT sebagai basis *classifier*. Beberapa aspek penting dalam metode RF di antaranya melakukan *bootstrap sampling* untuk membangun pohon prediksi, masing-masing DT memprediksi dengan prediktor acak dan RF sendiri melakukan prediksi dengan mengombinasikan hasil dari setiap DT dengan cara *majority vote* untuk klasifikasi dan juga rata-rata untuk regresi [17].

E. Evaluasi

Evaluasi model pada makalah ini menggunakan *confusion matrix*. *Confusion matrix* memberikan informasi berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi yang telah digunakan berdasarkan hasil prediksi dibandingkan dengan data yang benar. Proses klasifikasi *confusion matrix* menghasilkan empat nilai, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), seperti ditunjukkan pada Tabel I. Dengan mengetahui nilai TP, TN, FP, dan FN, dapat diperoleh nilai *precision*, *recall* dan *f1-score*.

Confusion matrix menampilkan tingkat akurasi dari proses klasifikasi yang telah dilakukan. *Precision* adalah proporsi dari pelabelan yang teridentifikasi dengan benar. Untuk menghitung nilai *precision*, digunakan (2).

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{2}$$

TABEL II
DATASET

Domain	Kalimat	Aspek
Restoran	1.521	makanan, layanan, harga, dan tempat
Hotel	2.880	AC, air panas, bau, <i>general</i> , kebersihan, handuk, pelayanan, sarapan, TV, dan Wi-Fi
<i>E-commerce</i>	572	produk, <i>packaging</i> , dan pengiriman

TABEL III
TEXT PREPROCESSING

	Sebelum	Sesudah
<i>Case Folding</i>	Suasananya sangat romantis dan menyenangkan apalagi jika datang dimalam hari selain suasananya, makanan disajikan ENAK dan VARIATIF.	suasananya sangat romantis dan menyenangkan apalagi jika datang dimalam hari selain suasananya makanan disajikan enak dan variatif
<i>Stopword Removal</i>	suasananya sangat romantis dan menyenangkan apalagi jika datang dimalam hari selain suasananya makanan disajikan enak dan variatif	suasananya romantis menyenangkan dimalam suasananya makanan disajikan variatif
<i>Stemming</i>	suasananya romantis menyenangkan dimalam suasananya makanan disajikan variatif	suasana romantis tenang malam suasana makan saji variatif

Recall adalah proporsi dari informasi yang dapat ditemukan dari pelabelan. Persamaan (3) merupakan rumus untuk menghitung nilai *recall*.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

Precision dan *recall* digunakan untuk mendapatkan nilai *f1-score*, seperti ditunjukkan pada (4).

$$F1 - score = \frac{2 \times recall \times precision}{recall + precision} \times 100\% \quad (4)$$

IV. PROSES PENGUJIAN

A. Dataset

Penelitian dimulai dengan mengumpulkan *dataset* dari tiga domain bahasa Indonesia, yaitu domain restoran, hotel, dan *e-commerce*. Ketiga domain tersebut sudah tersedia dan sudah berlabel untuk masing-masing aspeknya. Tabel II menunjukkan bahwa pada domain restoran terdapat empat aspek, yaitu makanan, layanan, harga, dan tempat, yang terdiri atas 1.521 kalimat.

Pada domain hotel terdapat sepuluh aspek, yaitu AC, air panas, bau, *general*, kebersihan, handuk, pelayanan, sarapan, TV, dan Wi-Fi, yang terdiri atas 2.880 kalimat, sedangkan domain *e-commerce* memiliki tiga aspek, yaitu produk, *packaging*, dan pengiriman, yang terdiri atas 572 kalimat.

TABEL IV
PERHITUNGAN TF, DF, IDF

Term	TF			DF	IDF
	D1	D2	D3		
Harga	2	0	0	2	0,1761
Tempat	1	1	0	2	0,1761
Enak	1	0	1	2	0,1761
Ramah	0	1	0	1	0,4771

TABEL V
HASIL TF- IDF

Term	TF		
	D1	D2	D3
Harga	0,3522	0,0000	0,0000
Tempat	0,1761	0,1761	0,0000
Enak	0,1761	0,0000	0,1761
Ramah	0,0000	0,4771	0,0000

B. Text Preprocessing

Tahapan *text preprocessing* terdiri atas tiga proses, yaitu *case folding*, *stopword removal*, dan *stemming*. Tabel III menunjukkan cara tiap-tiap proses dijalankan dengan menggunakan *library NLTK* dan *library sastrawi*. Penggunaan *case folding* bertujuan untuk mengubah kata yang terdapat pada ulasan menjadi karakter huruf kecil dan menghilangkan tanda baca, seperti pada ulasan “ENAK” yang menggunakan huruf besar menjadi “enak” dan menghilangkan tanda baca “,” dan “.”. Tahapan selanjutnya, *stopword removal*, berfungsi untuk menghapus kata yang tidak berkontribusi pada keakuratan hasil klasifikasi. Kata yang dihapus sudah terdapat dalam *list stopwords*, seperti “sangat”, “dan”, serta “jika”. Proses selanjutnya yaitu *stemming*, yang bertujuan untuk menghilangkan imbuhan, sisipan, dan akhiran, seperti pada kata “suasananya” menjadi “suasana”, “menenangkan” menjadi “tenang”, dan “disajikan” menjadi “saji”.

C. Term Weighting

Perhitungan *term weighting* dilakukan dengan menggunakan TF-IDF, sedangkan *Library TfidfVectorizer* digunakan untuk menampilkan bobot kata sesuai dengan kemunculan kata. Perhitungan IDF menghasilkan bobot yang sesuai dengan tingkat keunikan dari kata tersebut dalam indeks kata. Tabel IV menunjukkan contoh proses perhitungan TF, *Document Frequency* (DF), dan IDF. Nilai TF-IDF diperoleh melalui perhitungan dengan menggunakan (1). Hasil dari contoh perhitungan ditunjukkan pada Tabel V.

D. Pembagian Dataset

Dataset dari ketiga domain yang ada dibagi menjadi dua, yaitu data *training* dan data *testing*. Data *training* digunakan untuk analisis pola dan membuat model, sedangkan data *testing* digunakan untuk menguji model yang telah dibangun. *Dataset* restoran sudah dibagi menjadi *dataset training* dan *dataset testing* sesuai pada penelitian terdahulu [12]. Dua *dataset* lainnya dibagi menjadi 80% data *training* dan 20% data *testing*. Pembagian kedua *dataset* tersebut dilakukan secara acak



Gbr. 2 Hasil perbandingan.

TABEL VI
PEMBAGIAN DATASET

Domain	Aspek	Training	Testing
Restoran	Makanan	503	117
	Layanan	97	22
	Harga	253	29
	Tempat	440	137
Hotel	AC	471	121
	Air Panas	386	71
	Bau	371	85
	General	276	59
	Kebersihan	965	237
	Handuk	695	169
	Pelayanan	643	148
	Sarapan	174	47
	TV	217	51
Wi-Fi	335	99	
E-commerce	Produk	315	85
	Packaging	132	30
	Pengiriman	169	59

TABEL VII
HASIL PERHITUNGAN CONFUSION MATRIX

	TP	FP	FN	TN	Precision	Recall	F1-score
Makanan	111	51	6	254	0,685	0,949	0,796
Layanan	19	4	3	396	0,826	0,864	0,844
Harga	27	1	2	392	0,964	0,931	0,947
Tempat	124	57	13	228	0,685	0,905	0,780

dengan menggunakan *sklearn.model_selection library* sehingga pembagian *dataset e-commerce* sebanyak 457 kalimat untuk training dan 115 kalimat untuk testing. Sementara itu, pada *dataset hotel*, diperoleh 2.304 kalimat data *training* dan 576 kalimat data *testing*. Pembagian aspek di tiap domain ditunjukkan pada Tabel VI.

E. Klasifikasi

Tahapan klasifikasi aspek menggunakan tiga metode yang terdapat pada algoritme *machine learning*, yaitu NB, SVM, dan RF. Pada algoritme *machine learning*, metode yang dipilih dilatih untuk mengenali sebuah label yang terdapat pada aspek menggunakan *data training* yang ada. Metode NB menggunakan *MultinomialNB library*, metode SVM menggunakan *linearSVC library*, sedangkan metode RF menggunakan *RandomForestClassifier library* untuk membuat model *training*. Setelah mendapatkan model klasifikasi sesuai masing-masing metode, data *testing* yang tersedia dimasukkan

ke dalam model yang sudah dibuat untuk mendapatkan hasil prediksi.

F. Evaluasi

Hasil prediksi dari klasifikasi aspek lalu dibandingkan dengan data *testing* yang aktual sehingga didapatkan *confusion matrix* untuk masing-masing aspek. Berdasarkan nilai *confusion matrix* yang telah didapatkan, selanjutnya dicari nilai *precision*, *recall*, dan *f1-score* sesuai dengan perhitungan pada (2) sampai (4). Tabel VII menunjukkan hasil TP, FP, FN, dan TN dari masing-masing aspek yang terdapat pada *dataset restoran* dengan menggunakan algoritme RF. Kemudian, dari hasil masing-masing aspek tersebut, diambil rata-rata untuk mendapatkan nilai *precision*, *recall*, dan *f1-score* untuk *dataset restoran*.

Pengujian hasil klasifikasi aspek pada ketiga *dataset* tersebut menggunakan *sklearn.metrics library* untuk mendapatkan nilai *precision*, *recall*, dan *f1-score* pada masing-masing algoritme pada *machine learning*.

Berdasarkan perhitungan pada tahap evaluasi, didapatkan nilai *precision*, *recall*, dan *f1-score* pada tiap-tiap aspek yang terdapat pada tiga *dataset* yang berbeda, seperti ditunjukkan Tabel VIII, bahwa semua aspek memiliki nilai *precision*, *recall*, dan *f1-score* yang kemudian dicari rata-ratanya untuk mendapatkan nilai dari masing-masing domain dan metode yang digunakan.

V. HASIL

Hasil penelitian menggunakan tiga domain *dataset* berbahasa Indonesia dengan pendekatan tiga metode pada

TABEL VIII
HASIL UJI

	SVM			NB			RFC		
	P	R	F1	P	R	F1	P	R	F1
Restoran									
Makanan	0,70	0,90	0,78	0,64	0,90	0,75	0,69	0,95	0,80
Layanan	0,77	0,77	0,77	1,00	0,32	0,48	0,83	0,86	0,84
Harga	1,00	0,93	0,96	1,00	0,31	0,47	0,96	0,93	0,95
Tempat	0,70	0,82	0,76	0,65	0,81	0,72	0,69	0,91	0,78
Rata-rata	0,793	0,855	0,818	0,823	0,585	0,605	0,793	0,913	0,843
Hotel									
AC	0,96	0,89	0,93	1,00	0,35	0,52	0,96	0,98	0,97
Air panas	0,96	0,75	0,84	0,93	0,18	0,31	0,97	0,87	0,92
Bau	0,89	0,58	0,70	1,00	0,08	0,15	0,89	0,66	0,76
General	0,76	0,37	0,50	0,00	0,00	0,00	0,74	0,39	0,51
Kebersihan	0,94	0,82	0,88	0,86	0,64	0,74	0,89	0,90	0,90
Handuk	0,94	0,74	0,83	0,94	0,47	0,63	0,92	0,78	0,84
Pelayanan	0,91	0,74	0,81	0,91	0,41	0,57	0,91	0,85	0,88
Sarapan	0,97	0,64	0,77	0,00	0,00	0,00	0,98	0,85	0,91
TV	0,95	0,73	0,82	1,00	0,12	0,21	0,94	0,94	0,94
Wi-Fi	1,00	0,85	0,92	0,92	0,24	0,38	0,98	0,90	0,94
Rata-rata	0,928	0,711	0,800	0,756	0,249	0,351	0,918	0,812	0,857
E-commerce									
Produk	0,86	0,87	0,87	0,75	0,95	0,84	0,81	0,98	0,89
Packaging	0,94	0,50	0,65	1,00	0,37	0,54	1,00	0,83	0,91
Pengiriman	0,98	0,69	0,81	1,00	0,63	0,63	0,98	0,80	0,88
Rata-rata	0,927	0,687	0,777	0,917	0,650	0,670	0,930	0,870	0,893

Ket.: P = precision; R = recall; F1 = f1-score

TABEL IX
PERBANDINGAN DENGAN PENELITIAN TERDAHULU

Category	[12]			Pengujian yang Dilakukan		
	P	R	F1	P	R	F1
Makanan	0,658	0,906	0,763	0,661	0,932	0,773
Layanan	0,739	0,773	0,756	0,769	0,909	0,833
Harga	1,000	1,000	1,000	1,000	0,966	0,982
Tempat	0,705	0,860	0,775	0,683	0,898	0,776
Average	0,776	0,885	0,824	0,778	0,926	0,841

Ket.: P = precision; R = recall; F1 = f1-score

machine learning, yaitu NB, SVM dan RF, serta TF-IDF sebagai term weighting menghasilkan RF sebagai metode klasifikasi yang baik. Hal ini dapat dilihat pada Gbr. 2 yang menunjukkan hasil keseluruhan dari pengujian. Berdasarkan hasil perbandingan pada Gbr. 2, domain restoran dengan menggunakan algoritme RF menghasilkan nilai tertinggi, dengan precision 79,3%, recall 91,3%, dan f1-score 84,3%. Begitu juga dengan domain e-commerce, didapatkan RF sebagai yang terbaik, dengan precision 93,0%, recall 87,0%, dan f1-score 89,3%. Hasil berbeda didapatkan pada dataset hotel, dengan SVM menghasilkan nilai tertinggi pada precision, yaitu 92,8%, dan algoritme RF menghasilkan nilai tertinggi pada recall dan f1-score, yaitu 87,0% dan 89,3%.

Pengujian pada domain restoran dengan menggunakan algoritme RF memperoleh hasil yang lebih baik dibandingkan dengan penelitian sebelumnya [12]. Tabel IX menunjukkan perbaikan precision, recall, dan f1-score, dengan nilai precision 0,778, recall 0,926, dan f1-score 0,84.

VI. KESIMPULAN

Berdasarkan hasil analisis dan beberapa pengujian yang telah dilakukan, dapat disimpulkan bahwa dengan menerapkan beberapa model algoritme klasifikasi yang terdapat pada machine learning untuk menentukan nilai precision, recall, dan f1-score, algoritme RF menghasilkan nilai yang terbaik untuk keseluruhan domain, dengan nilai precision 79,3%, recall 91,3%, dan f1-score 84,3% pada domain restoran; nilai precision 93,0%, recall 87,0%, dan f1-score 89,3% pada domain e-commerce; serta nilai precision 91,8%, recall 81,2%, dan f1-score 85,7% pada domain hotel. Metode RF hanya kalah dengan metode SVM pada precision di domain e-commerce, dengan perbandingan nilai 92,8% dan 91,8%. Hasil penelitian yang dilakukan ini lebih baik daripada penelitian sebelumnya.

Masalah pada penelitian ini adalah masih adanya imbalance dataset yang digunakan sehingga berpengaruh pada nilai precision, recall, dan f1-score. Untuk mengatasi permasalahan ini, dapat digunakan oversampling atau resampling pada penelitian selanjutnya sehingga dapat meningkatkan nilai precision, recall, dan f1-score.

REFERENSI

- [1] D.F. Nasiri dan I. Budi, "Aspect Category Detection on Indonesian E-commerce Mobile Application Review," *Int. Conf. Data Softw. Eng.*, 2019, hal. 1–6.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*, Chicago, USA: Morgan and Claypool Publisher, 2012.
- [3] T. Alvarez-López, J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, dan F.J. González-Castã, "SVM and CRF for Aspect Detection and Unsupervised Aspect-Based Sentiment Analysis," *Proc. 10th Int. Work. Semant. Eval.*, 2016, hal. 306–311.
- [4] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, dan S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," *Proc. 8th Int. Work. Semant. Eval.*, 2015, hal. 27–35.
- [5] H. Papageorgiou, I. Androutsopoulos, D. Galanis, M. Pontiki, dan S. Manandhar, "SemEval-2015 Task 12: Aspect Based Sentiment Analysis," *Proc. 9th Int. Work. Semant. Eval.*, 2015, hal. 486–495.
- [6] M. Pontiki, D. Galanis, H. Papageorgiou, dkk., "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," *Proc. 10th Int. Work. Semant. Eval.*, 2016, hal. 19–30.
- [7] S. Movahedi, E. Ghadery, H. Faili, dan A. Shakery, "Aspect Category Detection via Topic-Attention Network," *arXiv Prepr. arXiv1901.01183*, hal. 1–9, 2019.
- [8] M. Afzaal, M. Usman, dan A. Fong, "Tourism Mobile App with Aspect-Based Sentiment Classification Framework for Tourist Reviews," *IEEE Trans. Consum. Electron.*, Vol. 65, No. 2, hal. 233–242, 2019.
- [9] M.S. Mubarak, A. Adiwijaya, dan M.D. Aldhi, "Aspect-based Sentiment Analysis to Review Products Using Naïve Bayes," *AIP Conf. Proc.*, Vol. 1867, No. 1, 2017, hal. 1–8.
- [10] A. Bhoi dan S. Joshi, "Various Approaches to Aspect-based Sentiment Analysis," *arXiv:1805.01984*, hal. 1-3, 2018.
- [11] S. Gojali dan M. L. Khodra, "Aspect Based Sentiment Analysis for Review Rating Prediction," *Int. Conf. Adv. Informatics Concepts, Theory Appl.*, 2016, hal. 1–6.
- [12] D. Ekawati dan M.L. Khodra, "Aspect-based Sentiment Analysis for Indonesian Restaurant Reviews," *Int. Conf. Adv. Informatics Concepts, Theory Appl.*, 2017, hal. 1–6.
- [13] A. Cahyadi dan M.L. Khodra, "Aspect-Based Sentiment Analysis Using Convolutional Neural Network and Bidirectional Long Short-Term Memory," *Int. Conf. Adv. Informatics Concept Theory Appl.*, 2018, hal. 124–129.
- [14] Z. Fachrina dan D.H. Widyantoro, "Aspect-Sentiment Classification in Opinion Mining using the Combination of Rule-Based and Machine Learning," *Int. Conf. Data Softw. Eng.*, 2017, hal. 1–6.
- [15] A. Ilmania, S. Cahyawijaya, Abdurrahman, dan A. Purwarianti, "Aspect Detection and Sentiment Classification Using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis," *Int. Conf. Asian Lang. Process.*, 2018, hal. 62–67.
- [16] Maulana, "Ringkasan Jumlah Aspek Ulasan Hotel untuk Pembentukan Dataset Sentimen Analisis Berbasis Aspek," *J. Linguist. Komputasional*, Vol. 3, No. 2, hal. 62–66, 2020.
- [17] M.A. Fauzi, "Random Forest Approach for Sentiment Analysis in Indonesian Language," *Indones. J. Electr. Eng. Comput. Sci.*, Vol. 12, No. 1, hal. 46–50, 2019.
- [18] (2018) GITHUB website. [Online], <https://github.com/PratamaAgung/AspectBasedSentimentAnalysis>, tanggal akses: 22-Jan-2020.
- [19] (2019) GITHUB website. [Online], <https://github.com/annisanurulazhar/absa-playground>, tanggal akses: 11-Feb-2021.
- [20] M.S. Akhtar, D. Gupta, A. Ekbal, dan P. Bhattacharyya, "Feature Selection and Ensemble Construction: A Two-step Method for Aspect Based Sentiment Analysis," *Knowledge-Based Syst.*, Vol. 125, hal. 116–135, 2017.
- [21] D. Tang, B. Qin, dan T. Liu, "Aspect Level Sentiment Classification with Deep Memory Network," *Conf. Empir. Methods Nat. Lang. Process.*, 2016, hal. 214–224.
- [22] A. Mridula dan C.R. Kavitha, "Opinion Mining and Sentiment Study of Tweets Polarity Using Machine Learning," *2018 2nd Int. Conf. Inven. Commun. Comput. Technol.*, 2018, hal. 621–626.
- [23] N.U. Pannala, C.P. Nawarathna, J.T.K. Jayakody, L. Rupasinghe, dan K. Krishnadeva, "Supervised Learning Based Approach to Aspect Based Sentiment Analysis," *Int. Conf. Comput. Inf. Technol.*, 2016, hal. 662–666.