Klasifikasi Review Customer *E-Commerce* Menggunakan Metode K-Nearest Neighbor (Studi Kasus: Bukalapak)

1st Shinta Pramuwidya
Fakultas Rekaya Industri
Universitas Telkom
Bandung, Indonesia
pramuwidyashinta@student.telkomuniversity.ac.id

2nd Riska Yanu Fa'rifah
Fakultas Rekaya Industri
Universitas Telkom
Bandung, Indonesia
riskayanu@telkomuniversity.ac.id

3rd Oktariani Nurul Pratiwi
Fakultas Rekaya Industri
Universitas Telkom
Bandung, Indonesia
onurulp@telkomunivesity.ac.id

Abstrak— Melihat persaingan *e-commerce* yang semakin ketat saat ini membuat Bukalapak melakukan berbagai upaya agar dapat bertahan serta meningkatkan kualitas layanan terhadap konsumen. Salah satu cara yang dapat dilakukan adalah melakukan evaluasi dari hasil *review*. Untuk dapat mengambil sebuah keputusan dari hasil *review* langkah yang perlu diambil salah satunya dengan mengklasifikasikan *review* dengan bertujuan untuk mengkategorikan data terhadap komentar atau *review* sehingga dapat membantu pelaku usaha dalam menarik kesimpulan terkait kecenderungan komentar. *Dataset* yang telah dikumpulkan dilakukan *preprocessing* sehingga berjumlah sebanyak 87.241 data. Di karenakan data memiliki *missing value* maka diatasi dengan metode imputasi menggunakan *mode()*. Setelah mengatasi *missing value*, *dataset* dihitung pembobotannya dengan tfidfVectorizer, selanjutnya di *resampling* dengan SMOTE agar data seimbang. *Review* dianalisis dengan algoritma K-Nearest Neighbors dengan tiga skenario yaitu 60:40, 70:30, dan 80:20, serta memiliki tiga jenis *k_neighbors* yaitu k=3, k=5 dan k=7. Jarak yang digunakan pada penelitian ini adalah Euclidean. Hasil analisis menunjukan bahwa KNN terbaik ada pada rasio training dan testing 80:20 dengan k=3. Hasil analisis menunjukan hasil evaluasi sebesar 76,42%. Hasil klasifikasi dengan KNN menunjukan komentar negatif lebih banyak daripada yang positif. Dari hasil penelitian ini diharapkan dapat dijadikan bahan evaluasi bagi Bukalapak untuk meningkatkan kualitas layanan.

Kata kunci— klasifikasi, K-Nearest neighbors, euclidean

I. PENDAHULUAN

A. Latar Belakang

Seiring berjalannya waktu, teknologi informasi pasti terus mengalami perkembangan inovasi dan bertransformasi ke arah yang semakin canggih. Hal itu dibuktikan dengan kehadiran teknologi informasi yang bergerak di bidang perdagangan seperti *electronic commerce (e-commerce). E-commerce* merupakan proses atau kegiatan bisnis transaksi yang melibatkan barang dan jasa menggunakan teknologi informasi [1]. Di Indonesia sudah banyak perusahaan *e-commerce* yang berkembang, salah satunya adalah Bukalapak. *E-commerce* yang berdiri pada 2010 ini merupakan salah satu *e-commerce* yang diciptakan oleh anak bangsa. Dalam perkembangannya, Bukalapak mampu bersaing dengan *e-commerce* yang didirikan oleh perusahaan asing, salah satunya adalah Lazada. Bukalapak berhasil menduduki urutan ketiga dalam top 10 *e-commerce* yang ada di Indonesia dan masih memiliki peluang untuk terus meningkat dalam aspek pelayanan terhadap *customer*. Data grafik tersebut dikeluarkan oleh Iprice Insight [2]. Hal ini akan menimbulkan persaingan dalam mekanisme pasar yang memacu Bukalapak untuk terus berinovasi menghasilkan produk serta pelayanan yang bervariatif dengan persaingan harga yang menguntungkan bagi pihak produsen maupun konsumen [3]. *Text mining* merupakan penambangan dokumen teks dari website yang berisi komentar, pendapat, ulasan, *feedback*, kritik, dan *review* yang menjadi hal penting, alasan *text mining* menjadi teknik atau metode untuk mendapatkan hasil *review* yang digunakan sebagai bahan evaluasi yaitu karena jumlah dari hasil *review* yang tidak sedikit serta berasal dari berbagai bentuk sehingga perlu adanya perapian data apabila data dikelola dengan baik maka dapat memberikan keuntungan berupa informasi yang bermanfaat untuk membantu individu atau organisasi dalam pengambilan sebuah keputusan [4]. Dalam proses pengklasifikasian terdapat target variabel kategori, dipisahkan dalam dua kategori yaitu :

a. Komentar bersifat negatif, bertujuan untuk mengetahui kritik dan saran yang diberikan konsumen untuk dijadikan bahan evaluasi bagi Bukalapak agar dapat memperbaiki pelayanan sehingga mampu menarik minat konsumen.
b. Komentar bersifat positif, bertujuan untuk mengetahui penilaian konsumen terkait kecenderungan minat jenis produk atau layanan [5]

Berdasarkan deskripsi di atas penelitian ini akan melakukan analisis klasifikasi hasil *review customer* Bukalapak dengan menggunakan algoritma K-Nearest Neighbor (KNN) untuk mendapatkan hasil akurasi yang lebih baik karena melihat jarak antara objek dengan kelas untuk mengetahui kecenderungan ketegori komentar guna meningkatkan layanan.

II. KAJIAN TEORI

A. *E-Commerce*

E-commerce adalah pembelian atau penjualan barang antara bisnis, rumah tangga, individu, pemerintah, dan organisasi publik serta swasta lainnya melalui jaringan komputer. Definisi sempit di sisi lain hampir sama dengan definisi luas kecuali instrumen perdagangan terbatas dengan internet. Kerangka utama dari e-commerce terdiri dari *people* (penjual, pembeli, perantara, sistem informasi dan lainnya), *public policy* (kebijakan dan peraturan publik seperti pajak,

regulasi dan lainnya), *marketing* dan *advertising* (pemasaran periklanan seperti promosi, konten web, target pemasaran dan lainnya), *support service* (layanan pendukung seperti logistik, pembayaran, keamana sistem dan jaringan dan lainnya), *business partnership* (kemitraan bisnis seperti program afiliasi, pertukaran dan lainnya) [6].

B. *Preprocessing*

Secara umum proses yang dilakukan dalam tahapan preprocessing adalah sebagai berikut:

1. Pembersihan data
   Pembersihan data merupakan proses pembersihan kata dengan menghilangkan delimiter koma (,), titik (.), dan tanda baca lainnya serta menghilangkan *backslices*. Selain menghilangkan delimiter tanda baca, pembersihan data juga merupakan proses menghapus duplikat, menghapus suku kata yang berulang dan menghapus angka dalam satu kalimat [7].
2. *Case Folding*
   Case folding merupakan tahapan awal pada *preprocessing* yang bertujuan untuk mengubah setiap bentuk kata menjadi sama. Tidak semua dokumen teks konsisten dalam penggunaan huruf capital [8].
3. *Slang words*
   Slang words atau yang dapat disebut normalisasi bahasa merupakan tahapan yang dilakukan untuk mengembalikan bentuk penulisan dari masing-masing kata yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). [7].
4. *Tokenizing*
   Tokenizing merupakan proses memacah kalimat menjadi kata-kata yang dilakukan untuk menjadi sebuah kalimat yang lebih bermakna [7].
5. *Stopword Removal*
   Stopword removal merupakan tahap untuk menghilangkan kata umum yang tidak memiliki arti penting dan tidak digunakan namun seringkali muncul dalam dokumen. [7].
6. *Stemming*
   Stemming merupakan tahap terakhir dalam proses *preprocessing*. Proses yang dilakukan untuk mencari *stem* (kata dasar) dari kata hasil *stopword removal* (filtering) dengan menggunakan *library* Sastrawi untuk proses *stemming* bahasa Indonesia [7].

C. Klasifikasi

Klasifikasi merupakan salah satu tugas yang penting *text mining*. Sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan. Klasifikasi merupakan pengelompokan fitur ke dalam kelas yang sesuai. Seperti yang telah dinyatakan sejumlah klasifikasi teknik telah diusulkan dalam literatur [9].

D. *K-Nearest neighbors*

Algoritma K-Nearest Neighbor (KNN) termasuk dalam *supervised learning*, dimana hasil *query instance* yang baru diklasifikasi berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN [10] artinya melakukan klasifikasi pada objek berdasarkan data yang jaraknya peling dekat dengan objek tersebut [11]. Metode ini bertujuan untuk mengklasifikasi Objek baru berdasarkan atribut dan *training* sample [12].

Adapun tahapan yang dilakukan untuk menghitung metode K-Nearest Neighbor (KNN) antara lain:

1. Menentukan jumlah tetangga *k*
2. Menghitung jarak objek dengan masing-masing data kelompok yang terdiri dari data latih dan data uji. Perhitungan jarak menggunakan rumus Euclidean *distance*.
3. Kemudian didapatkan hasil dari perhitungan pengklasifikasian [13].

Rumus jarak Euclidean didefinisikan dalam persamaan:

$$d_i = \sqrt{\sum_{k=1}^{n} (p_i - q_i)^2}$$

Keterangan:

d_i = Jarak variabel data
p_i = Sample data
q_i = Data uji
i = Variable data
n = Dimensi data

Setelah semua langkah dilakukan pengklasifikasi menggunakan KNN, langkah selanjutnya adalah melakukan validasi untuk mencocokan dengan data pakar dan melihat nilai akurasinya [13].

E. Confusion Matrix

Confusion Matrix merupakan sebuah tabel yang terdiri atas banyaknya baris data uji yang diprediksi benar atau tidak benar oleh model klasifikasi, tabel ini diperlukan untuk menentukan kinerja suatu model klasifikasi. Maka dari itu metode ini cocok untuk digunakan dalam penelitian ini guna mengukur seberapa akurat hasil klasifikasi dari model yang telah dibuat [14].

Variabel yang digunakan pada tabel didapat dari confusion matrix, dimana TP (*true positive)*, TN (*true negative)*, FP (*false negative)*, FN (*false negative)*. Berikut tabel confusion Matrix [15]:

TABEL II-1
CONFUSION MATRIX

Confusion Matrix		*Predicted*	
		Positive	*Negative*
Actual	*Positive*	TP	FN
	Negative	FP	TN

F. *K-Fold Cross Validation*

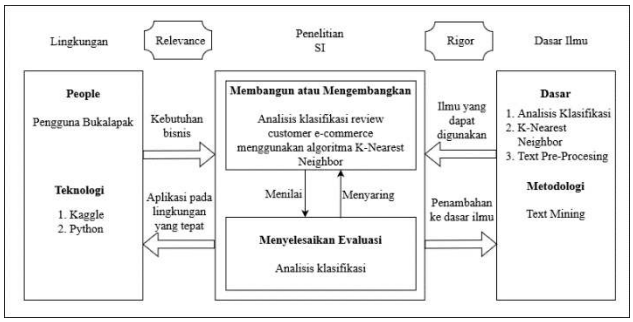
K-fold cross validation adalah salah satu dari jenis pengujian *cross validation* yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokan data tersebut sebanyak nilai k k-fold. Kemudian salah satu kelompok k-fold tersebut akan dijadikan sebagai data uji sedangkan sisa kelompok yang lain akan dijadikan data latih [16].

III. METODE

A. Model Konseptual

Model konseptual merupakan ilustrasi yang menggambarkan sebuah konsep pemikiran untuk menjelaskan sebuah pemecah masalah yang dapat dihasilkan dari aspek teoritis

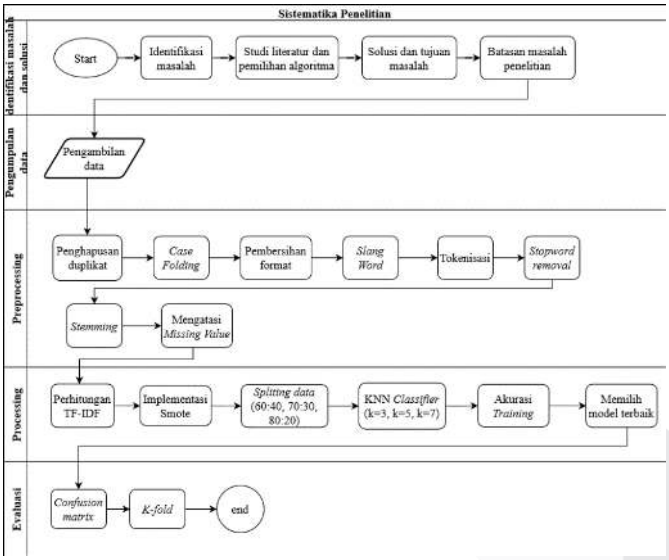
dan aspek hipotesis serta bisa melakukan simulasi dari kebutuhan [17].



GAMBAR III-1
MODEL KONSEPTUAL

Berdasarkan gambar di atas, menggambarkan bahwa lingkungan pada aspek *people* peneliti menggunakan komentar/*review* pengguna Bukalapak yang bersumber dari beberapa media sosial, kemudian pada aspek teknologi (*tools)* yang digunakan adalah kaggle sebagai sumber dataset dan bahasa pemrograman *python* untuk mengolah data. Dengan metodologi yaitu *text mining* melalui tahap *preprocessing* dan *processing* kemudian tahap selanjutnya analisis klasifikasi dengan algoritma K-Nearest Neighbor.

B. Sistematika Penyelesaian Masalah



GAMBAR III-2
SISTEMATIKA PENYELESAIAN MASALAH

Pada sistematika penyelesaian masalah terbagi menjadi lima tahap yaitu tahap Identifikasi masalah dan solusi, pengumpulan data, *preprocessing*, *processing*, dan pengambilan kesimpulan serta saran. Pada tahap identifikasi masalah, yang dilakukan pertama kali adalah melihat ulasan/*review* pengguna pada beberapa aplikasi *e-commerce* kemudian melakukan studi literatur untuk menentukan studi kasus dan memilih algoritma yang tepat untuk menentukan solusi yaitu dilakukannya sebuah penelitian dan menentukan tujuan dari permasalahan terakhir menentukan batasan masalah penelitian agar permasalahan yang diteliti tidak meluas cakupannya. Tahap kedua yaitu pengumpulan data, pengumpulan data dilakukan dengan mengambil dataset dari website kaggle.com . Tahap ketiga yaitu *preprocessing,* terdiri dari *pembersihan format, menghapus duplikat, case folding, slang word, tokenization, stopwords removal,* dan *steaming* selanjutnya data yang sudah bersih akan di periksa dan mengatasi *missing value* untuk hasil akurasi yang optimal dengan teknik imputasi atau mengisi nilai yang hilang dengan nilai yang yang didapat dari nilai-nilai yang diketahui salah satunya yaitu modus [18]. Selanjutnya tahap *processing*, pada tahap ini menghitung TF-IDF yang bertujuan untuk menghitungan kata umum yang ada pada information retrieval [7], selanjutnya melakukan *resampling* dengan metode *oversampling* yaitu SMOTE. Setelah itu melakukan *spliting data* atau pembagaian data menjadi data *training* dan data *testing* dengan tiga rasio 60:40, 70:30, 80:20 kemudian melakukan analisis klasifikasi dengan KNN Classifier tiga jenis k yaitu k=3, k=5, k=7 terdiri dari menentukan *k-neighbor* selanjutnya menghitung jarak Euclidean dengan formula pada Microsoft Excel kemudian mencari hasil akurasi dari *training data* terbaik untuk di evaluasi dengan *confusion matrix* untuk melihat nilai akurasinya kemudian di validasi dengan 10-*fold cross validation*.

C. Pengumpulan Data

Pada penelitian ini, Metode pengumpulan data yaitu dengan mengambil dataset dari website kaggle.com terkait komentar di akun Bukalapak yang telah disediakan sesuai dengan studi kasus yang diteliti. Dataset yang diambil adalah data yang bersifat data sekunder dengan jenis data yaitu data kualitatif. Setelah dataset terkumpul maka langkah selanjutnya mengurangi duplikat ulasan yang tidak relevan dengan studi kasus. Adapun pelabelan yang dilakukan dengan cara otomatis yaitu satu (positif) dan nol (negatif).

D. Pengolahan Data

Setelah melakukan pengumpulan data, data mentah akan dirapikan sebelum diolah untuk kemudian dianalisis mencari temuan-temuan sesuai dengan tujuan penelitian dengan 2 tahap yaitu *preprocessing* dan *processing*.

1. *Preprocessing*

   Preprocessing merupakan tahap ketiga setelah pengumpulan *dataset*. *Dataset* yang terkumpul akan dilakukan pembersihan data, *case folding*, *Slang words*, *tokenization*, *Stopwords removal*, *Stemming* terakhir memeriksa dan mengatasi missing data untuk hasil akurasi yang optimal [19].

TABEL III-1
HASIL *PREPROCESSING*

Proses	Hasil Preprocessing
Cleansing data	barang dengan kode pemesanan uda di terima thanks ya
Case folding	pengiriman cepat n barang bgus
Slang word	kecil tidak muat di pakai dan tidak bisa melar
Tokeniza-tion	['pengiriman', 'cepat', 'dan', 'barang', 'bagus']
Stopword removal	['pengiriman', 'cepat', S'barang', 'bagus']
Stemming	kirim cepat barang bagus

2. Processing

   Processing merupakan tahap keempat yaitu kemudian menghitung TF-IDF

TABEL III-2
CONTOH PEMBOBOTAN TF-IDF

Review	TF		IDF	TF-IDF
	1	2		
barang	1	0	1.405465	0.534046
expired	1	0	1.405465	0.534046

selanjutnya melakukan *resampling* dengan metode *oversampling* yaitu SMOTE, *spliting data* menjadi data *training* dan data *testing* dengan tiga rasio 60:40, 70:30, 80:20,

TABEL III-3
TOTAL *SPLITTING DATA* SETELAH SMOTE

Rasio	Proses split data		Total data
	Data Training	Data Testing	
60:40	90.702	60.468	151.170
70:30	105.819	45.351	
80:20	120.936	30.234	

melakukan analisis dengan KNN *Classifier* tiga jenis k yaitu k=3, k=5, k=7, Mencari penggunaan k yang hasilnya terbaik kemudian menghitung jarak Euclidean dengan formula pada Microsoft Excel, mencari hasil akurasi dari *training data* terbaik untuk di evaluasi dengan *confusion matrix* untuk melihat nilai akurasinya kemudian di validasi dengan 10-*fold cross validation*.

E. Metode Evaluasi

Metode evaluasi dilakukan dengan menggunakan *confusion matrix* atau dapat disebut juga dengan *error matrix*, dengan menggunakan akurasi, *precision*, *recall* dan *f1-score* berdasarkan *confusion matrix* yang dihasilkan dari data testing. Tujuan dari *confusion matrix* yaitu untuk memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui [20].

IV. HASIL DAN PEMBAHASAN

A. Klasifikasi K-Nearest Neighbor

Untuk implementasi algoritma atau model yang dipilih yaitu algoritma K-Nearest Neighbor dengan menggunakan *library* yang tersedia pada bahasa pemrograman *python* untuk melakukan klasifikasi dengan melatih model. Berikut merupakan tabel akurasi pada data latih (*training*) pada setiap skenario. Berikut merupakan tabel akurasi *training* setiap skenario :

TABEL IV-1
HASIL AKURASI *TRAINING DATA*

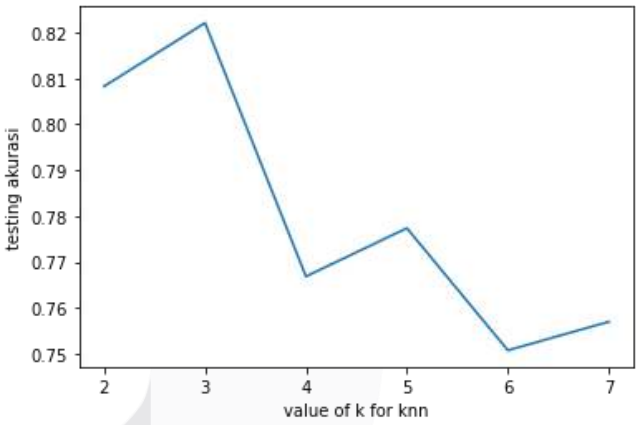
Rasio	Akurasi *Training*		
	n_neighbor= 3	n_neighbor = 5	n_ neighbor = 7
60:40	80,96	76,96	74,97
70:30	81,54	77,19	75,21
80:20	82,20	77,73	75,69

Dapat disimpulkan dari ketiga k maka akurasi tertinggi jatuh pada k=3 dengan rasio 80:20 dengan perolehan akurasi sebesar 82,20%.

TABEL IV-2
HASIL *EUCLIDEAN DISTANCE*

X (TF-IDF)	Label	Jarak	k=3	k=5	k=7
0,8371	0	0	0	0	0
0,5200	1	0,317			1
0,0652	1	0,772			
0,5704	0	0,267		0	0
0,8329	1	0,004	1	1	1
0,8754	1	0,038	1	1	1
0,4627	0	0,374			0
0,3908	1	0,446			
0,1528	0	0,684			
0,7191	0	0,118		0	0

Perhitungan jarak dilakukan dengan memasukan nilai TF-IDF dari sepuluh sampel kata untuk kemudian dihitung dengan persamaan jarak Euclidean.



GAMBAR IV-1
NILAI K UNTUK KNN

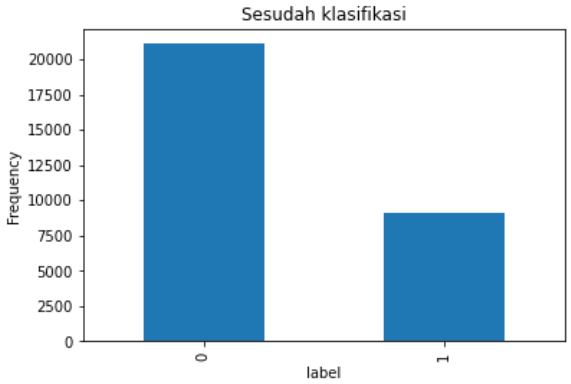
Berdasarkan tabel IV-3 akurasi *training* tertinggi diraih oleh k=3 pada skenario 60:40, 70:30, dan 80:20. Sehingga dapat disimpulkan dari ketiga k yang telah diuji berdasarkan akurasi *training* tertinggi jatuh pada k=3.

B. Evaluasi Performasi

Setelah melakukan pengujian model, penulis melakukan evaluasi terhadap model untuk mengetahui *precision*, *recall* dan *f1-score* dari rasio perbandingan 80:20 dan k=3 dan memvalidasi model dengan 10-*fold cross validation*.

TABEL IV-3
HASIL AKURASI *TESTING DATA*

Accuracy	Precision	Recall	F1-score
76%	93%	57%	71%



GAMBAR IV-2
TOTAL LABEL SESUDAH KLASIFIKASI

Setelah melalui proses testing maka didapatkan akurasi sebesar 76% dengan total komentar yang memiliki label negatif sebanyak 21.080 data dan total komentar yang memiliki label positif sebanyak 9.153 data.

TABEL IV-4
HASIL K-*FOLD CROSS VALIDATION*

K	Akurasi	Rata-Rata
1	0.703	70,57%
2	0.700	
3	0.708	
4	0.708	
5	0.702	
6	0.704	
7	0.709	
8	0.702	
9	0.708	
10	0.708	

Berdasarkan pengujian dari *K-fold cross validation* yang telah dilakukan maka didapatkan dari 10 *fold* yang memperoleh nilai tertinggi ada pada *fold* ke 7 dengan akurasi sebesar 70,95% , nilai terendah ada pada *fold* ke 2 dengan akurasi sebesar 70% dan memperoleh nilai rata-rata sebesar 70,57%.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut:

1. Klasifikasi terhadap *review customer* Bukalapak menggunakan algoritma K-Nearest Neighbors dilakukan beberapa tahap mulai dari preprocessing, ekstrasi fitur dengan TF-IDF dan klasifikasi menggunakan K-Nearest Neighbors. Akurasi hasil klasifikasi yang optimal didapatkan pada rasio 80:20 dengan k=3 kemudian perhitungan jarak menggunakan Euclidean.
2. Berdasarkan hasil evaluasi tertinggi dari proses *training*, maka tingkat akurasi algoritma KNN terhadap klasifikasi *review customer* Bukalapak untuk evaluasi proses testing dari skenario 80:20 dengan k=3 memiliki akurasi sebesar 76%, *precision* sebesar 93%, *recall* sebesar 57%, dan *f1-score* sebesar 71%. Perhitungan akurasi dan hasil evaluasi performasi mengunakan *confusion matrix* dan *k-fold cross validation*.
3. Berdasarkan hasil klasifikasi terhadap *review customer* Bukalapak maka menghasilkan sebanyak 21.080 data berasal dari komentar negatif dan 9.153 data berasal dari komentar positif. Maka dapat diketahui jika *review* komentar negatif lebih banyak dari komentar positif.

B. Saran

Saran yang dapat diberikan pada penelitian ini adalah sebagai berikut:

1. Diharapkan penelitian selanjutnya dapat mengatasi kata yang tidak baku atau singkatan-singkatan dengan menambah lebih banyak kamus kata pada kamus bahasa Indonesia guna mempermudah proses klasifikasi.
2. Diharapkan penelitian selanjutnya dapat mencari tahu cara agar nilai dari rata-rata *k-fold cross validation* tidak jauh berbeda dengan hasil akurasi *testing data* dengan metode *spliting data*.
3. Diharapkan penelitian selanjutnya dapat mengklasifikasi komentar berdasarkan jenis barang yang tersedia dalam *dataset*.

REFERENSI

[1] W. Febriantoro, "Kajian dan Strategi Pendukung Perkembangan E-Commerce Bagi UMKM di Indonesia," *Jurnal Manajemen dan Sistem Informasi,* p. 3, 2018.

[2] Iprice, "The Map of E-commerce in Indonesia," 16 November 2021. [Online]. Available: https://iprice.co.id/insights/mapofecommerce/en/.

[3] A. N. Hayati, "Analisis Tantangan dan Penegak Hukum Persaingan Usaha Pada Sektor E-commerce di Indonesia," *Jurnal Penelitian De Jure,* p. 110, 2021.

[4] S. Budi, "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means," *Jurnal Teknologi Informasi,* p. 1, 2017.

[5] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika,* pp. 215-216, 2016.

[6] D. Hendrasyah, "E-COMMERCE DI ERA INDUSTRI 4.0 DAN SOCIETY 5.0," *Jurnal Ilmiah Ekonomi Kita,* p. 2, 2019.

[7] Luqyana, W. Athira, I. Cholissodin and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* 2018.

[8] A. W. Luqyana, I. Cholissodin and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* p. 4705, 2018.

[9] A. P. Wibawa, M. G. A. Purnama, M. F. Akbar and F. A. Dwiyanti, "Metode-metode Klasifikasi," *Seminar*

Ilmu Komputer dan Teknologi Informasi, pp. 134-135, 2018.

[10] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indonesia Journal of Data and Science,* p. 30, 2020.

[11] A. A. T. Mara, E. Sediyono and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Journal Of Informatics Engineering,* 2021.

[12] J. J. A. Limbong, I. Sembiring and K. D. Hartomo, "ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR," *Jurnal Teknologi Informasi dan Ilmu Komputer,* 2021.

[13] Y. I. Claudy, R. S. Perdana and M. A. Fauzi, "Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (KNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* 2018.

[14] Bachtiar, M. Rivki and A. Mukharil, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR DALAM PENGKLASIFIKASIAN FOLLOWER TWITTER YANG MENGGUNAKAN BAHASA INDONESIA," *Jurnal Sistem Informasi,* 2017.

[15] S. Watmah, Suryanto and Martias, "Komparasi Metode K-NN, Support Vector Machine Dan Random Forest Pada E-Commerce Shopee," *Jurnal Inovasi dan Sains Teknik Elektro,* 2021.

[16] A. Hutapea, M. Furqon and Indriati, "Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* 2018.

[17] S. Robinson, G. Arbez, L. G. Birta, A. Tolk and G. Wagner, "Conceptual Modelling; Definition, Purpose And Benefits," *IEEE,* 2015.

[18] Susanti, S. Martha and E. Sulistianingsih, "K NEAREST NEIGHBOR DALAM IMPUTASI MISSING DATA," *Buletin Ilmiah Math. Stat dan Terapannya (BIMASTER),* 2018.

[19] C. Sukmayadi, A. Primajaya and I. Maulana, "Penerapan Algoritma K-Medoids dalam Menentukan Daerah Rawan Banjir di Kabupaten Karawang," *Jurnal Informatika,* 2021.

[20] K. S. Nugroho, "Confusion Matrix untuk Evaluasi Model pada Supervised Learning," 13 November 2019. [Online]. Available: https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f.