Faktor Exacta

Vol. 13, No. 4, Dec 2020, pp. 200~207
eISSN: 2502-339X, pISSN: 1979-276X, DOI: 10.30998/faktorexacta.v13i4.7074

Analisis Sentimen Pengguna *Marketplace* Bukalapak dan Tokopedia di *Twitter* Menggunakan *Machine Learning*

Irwansyah Saputra[1], Hanafi Eko Darono[2], Fachri Amsury[3], Muhammad Rizki Fahdia[4], Benni Ramadhan[5], Anggie Ardiansyah[6]
[1,2,3,4,5,6] STMIK Nusa Mandiri, Indonesia

Article Info

Article history:

Received Sep 9, 2019
Revised May 20, 2020
Accepted December 27, 2020

Keywords:

marketplace
twitter
data mining
rapidminer
sentiment analysis
algorithms

ABSTRACT

A collection of tweets from Twitter users about Marketplace Bukalapak and Tokopedia can be used as a sentiment analysis. The data obtained is processed using data mining techniques, in which there is a process of mining the text, tokenize, transformation, classification, stem, etc. Then calculated into three different algorithms to be compared, the algorithm used is the Decision Tree, K-NN, and Naïve Bayes Classifier with the aim of finding the best accuracy. Rapidminer application is also used to facilitate writers in processing data. The highest results from this study are Decision Tree algorithm with 82% accuracy, 81.95% precision and 86% recall.

Copyright © 2020 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Irwansyah Saputra,
STMIK Nusa Mandiri.
Email: irwansyah.iys@nusamandiri.ac.id

1. PENDAHULUAN

Saat ini sebagian besar pengguna internet dan media sosial telah mengunjungi toko online atau Marketplace. Bukalapak dan Tokopedia adalah Marketplace yang sering dikunjungi oleh pengguna media sosial Indonesia. Sementara itu salah satu media sosial yang paling sering digunakan dalam mengungkapkan pendapat dalam hal transaksi di Bukalapak dan Tokopedia adalah Twitter. Twitter adalah media sosial yang memungkinkan pengguna untuk mengirim pesan secara real-time.

Pendapat pelanggan melalui *Twitter* dapat menyimpulkan situs *online Marketplace* mana yang memiliki layanan terbaik. Ini karena banyaknya komentar yang ada di *Twitter* dan bahkan sering kali ada topik yang menjadi *trending* di *Twitter* tentang layanan yang disediakan oleh *Marketplace* [1]. Karena hal itulah penulis merasa perlu untuk menganalisa pandangan dan sentimen pengguna mengenai *Marketplace* ini.

Di sisi lain, data *mining* dapat diartikan sebagai metode yang digunakan untuk menemukan pengetahuan dari sebuah *database* atau untuk menemukan pengetahuan yang berguna dari data. Ini adalah proses untuk mendapatkan desain dan hubungan yang menarik sehingga data dapat dilayani dalam volume yang besar [2]. Dengan demikian data *mining* dapat digunakan sebagai analisis sentimen karena dapat mengolah data *tweet* dalam jumlah banyak.

Penelitian ini membahas bagaimana melakukan analisis sentimen mengguakan opini pelanggan *Marketplace* di *Twitter*, nantinya opini tersebut diklasifikasikan menjadi data opini sentimen positif dan sentimen negatif. Penelitian ini menggunakan tiga algoritma berbeda untuk mengekstrak pemikiran atau perasaan pengguna melalui *tweet* mereka dan mengelompokkan datanya ke dalam kategori yang berbeda. Kemudian membandingkan hasil dari ketiga algoritma tersebut untuk mengetahui *classifier* mana yang memberikan hasil terbaik dalam hal *accuracy*, *precision*, *dan recall* [3].

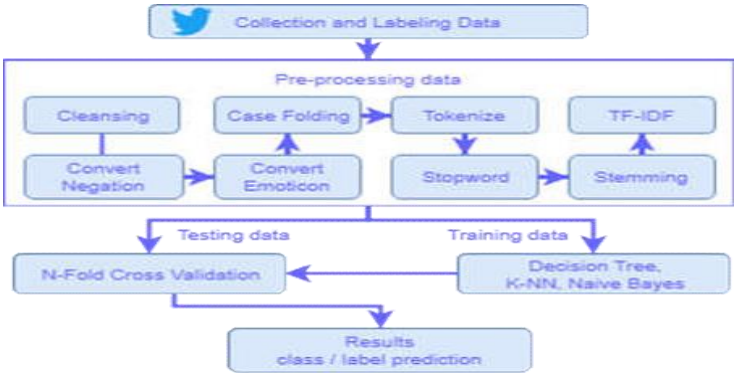
2. METODE

Evaluasi layanan situs online telah dilakukan. Penelitian tentang sentimen analisis dari tweet di Twitter menggunakan bahasa Arab untuk menganalisis akurasi dan memprediksi sentimen yang benar [4]. Penelitian telah mengusulkan sentimen analisis dari tweet bahasa Inggris menggunakan Rapidminer. Pendekatan alat digunakan untuk analisis sentimen tujuan bisnis. Penelitian lain menggunakan data Twitter untuk melakukan analisis linguistik dan kemudian membangun pengklasifikasi yang sangat efisien. Apa yang penulis papakarkan tadi adalah bukti bahwa sebelumnya sudah ada penelitian mengenai analisis sentimen dari penggunaan media sosial Twitter.

Analisis Sentimen dari *Twitter* harus fokus pada masalah klasifikasi [3]. Klasifikasi adalah proses menemukan model atau fungsi yang dapat menjelaskan dan membedakan suatu konsep atau kelas data [5]. Berbagai pendekatan klasifikasi seperti *Naïve Bayes*, *Support Vector Machine* (SVM), *dan K-Nearest Neighbor* (K-NN) telah diterapkan untuk menemukan hasil terbaik. Sebagai contoh, analisis klasifikasi sentimen dari media sosial *Twitter* telah dilakukan oleh F. Laeeq dan N. M. Tabrez. Dalam penelitian mereka, mereka menggunakan tiga pengklasifikasi yaitu K-NN, *Naïve Bayes* dan *Decision Tree Classifier* untuk mengklasifikasi sentimen dan memperoleh hasil yang menunjukkan akurasi K-NN, *Naïve Bayes*, dan *Decision Tree Classifier* masing-masing 77,50%, 80%, dan 78% [6].

Dengan demikian, penambangan dapat digunakan untuk tingkat analisis sentimen. Data *mining* adalah proses menemukan pola atau pengetahuan. Pola harus valid, berguna, dan dapat dipahami dalam sejumlah langkah: pra-pemrosesan, penambangan data, dan pasca-pemrosesan [3].

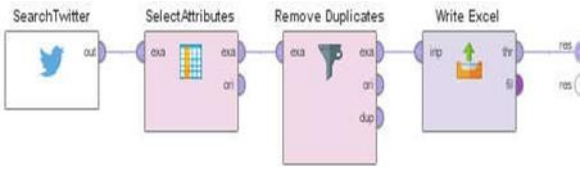
Dalam menganalisis sentimen, ada beberapa tahapan yang perlu dilakukan untuk mendapatkan hasil tes terbaik. Langkah-langkahnya terdiri dari pengumpulan dan pelabelan data, pra-pemrosesan, dan analisis Sentimen, Gambar 1 menunjukkan langkah-langkah dalam klasifikasi sentimen analisis yang diusulkan.



Gambar 1. Diagram Sistem Klasifikasi Analisis Sentimen

A. Pengumpulan dan Pelabelan Data

Tahap pertama dalam melakukan proses analisis sentimen adalah pengumpulan data. Data diambil dari *Twitter* dengan permintaan pencarian tentang Tokopedia dan Bukalapak masing-masing sebanyak 50 catatan menggunakan aplikasi *Rapidminer*.



Gambar 2. Pengambilan Data dari *Twitter*

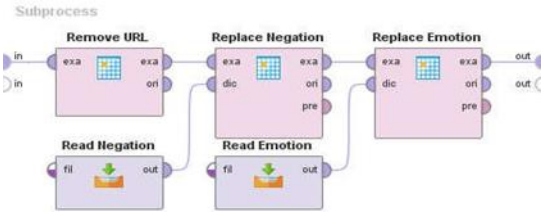
Gambar 2 menunjukkan pengambilan data dari *Twitter* menggunakan konten "*Search Twitter*" dan menyimpan data dalam file *excel* menggunakan konten "*Write Excel*" dengan hanya mengambil teks dari tweet dan menghapus semua *tweet* duplikat denga konten "*Remove Duplicates*". Tahap selanjutnya adalah memberi label. Pelabelan dilakukan untuk membagi data menjadi beberapa kelas sentimen yang akan digunakan. Jumlah kelas sentimen yang digunakan adalah dua kelas, yaitu negatif dan positif. Tujuan dari proses pelabelan ini adalah untuk membagi dataset menjadi 2 bagian, untuk menjadi data *training* dan data *testing*. Data *training* adalah data yang digunakan untuk melatih sistem agar dapat mengenali pola yang dicari, sedangkan data *testing* adalah data yang digunakan untuk menguji hasil pelatihan yang telah dilakukan [7]. Berikut ini adalah salah satu contoh *dataset* yang telah diberi label:

Text	Sentiment
@Bukalapak @Bukalapak tolong ditanggapi saya kecewa dengn bukalapak	Negative
Min sampe sekarang belum ada tindakan apapun dari admin Tokped ataupun penjual. Saya tidak mau mengulur2 waktu nih	Negative
btw, pas ngetik keywords 'sedotan' di tokopedia, hasil paling atas 'sedotan stainless'. wah, senang. sudah banyak yg makin peduli :)	Positive
Bukalapak Bikin Layanan Pinjaman Tanpa Syarat untuk UKM http://dlvr.it/Qb83n8 pic.twitter.com/8MucIMhUMO	Positive

Tabel 1. Contoh Label Dataset

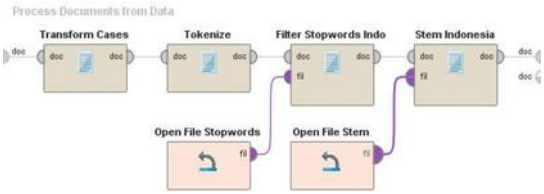
B. Pra – Pemrosesan

Setelah data diberi label, langkah selanjutnya adalah pra-pemrosesan. Tahap ini adalah tahap di mana data dipersiapkan untuk menjadi data yang siap dianalisis. Ada beberapa tahapan dalam preprocessing ini, termasuk *cleansing*, *convert negation*, *convert emoticons*, *case folding*, *tokenization*, *filtering stopword* dan *stemming* dalam bahasa Indonesia [3]. Gambar 3 menunjukkan operator "Subprocess". Dalam hal ini menggunakan konten "Remove URL", "Convert Negation", "Convert Emoticons".



Gambar 3. Operator *Subprocess*

Gambar 4 menunjukkan operator "*Process Document*", menggunakan konten "*Transform Cases*", "*Tokenize*", "*Stopword Filter*" dan "*Stem*" dalam bahasa Indonesia.



Gambar 4. Operator *Process Document*

Berikut ini adalah deskripsi terperinci dari tahapan pra-pemrosesan di atas:

1. *Cleansing*

Cleansing adalah tahap dimana karakter dan tanda baca yang tidak diperlukan dihapus dari teks. Bekerja untuk mengurangi *noise* pada dataset. Contoh karakter yang dihilangkan seperti URL, tag (#), tanda baca seperti titik (.), Koma (,) dan tanda baca lainnya [8]. Ini adalah contoh kalimat pembersihan data, *input* "Bukalapak Bikin Layanan Pinjaman Tanpa Ketentuan untuk UKM http://dlvr.it/Qb83n8pic.twitter.com/8MucIMhUMO", *output* "Bukalapak Bikin Layanan Pinjaman Tanpa Syarat untuk UKM".

2. *Convert Negation*

Dalam bahasa Indonesia, ada yang disebut kata negasi, yaitu kata yang dapat membalikkan makna kata yang sebenarnya [8]. Ini adalah contoh negasi konversi kalimat, *input* "Saya tidak mau mengulur2 waktu nih", keluaran "Saya tidak_mau mengulur2 waktu nih".

3. *Convert Emoticon*

Emosi adalah ekspresi wajah yang diwakili oleh kombinasi huruf, tanda baca, dan angka. Pengguna biasanya menggunakan emotikon untuk mengekspresikan perasaan mereka. *Convert emoticons* adalah salah satu cara untuk mengekspresikan ekspresi perasaan secara tekstual [8]. Ini adalah contoh kalimat *convert emoticon*, *input* "wah, senang. sudah banyak yang makin peduli :)", maka *output*nya "wah, senang. sudah banyak yang makin peduli senang". Berikut adalah salah satu contoh kata *convert emoticon*:

Sebelum	Sesudah
:(:'(:[;(:/): x(--" :# :-@ :c :f ;(:v :x :s)": *_*	sedih
:) :] (^_^) ^^v <3 ^^ ^_^ 0:) <:} :* (^.^) =) :3	senang

Tabel 2. Contoh *Convert Emoticon*

4. *Case Folding*

Dalam menulis *tweet*, biasanya ada bentuk huruf yang berbeda, tahap ini adalah proses keseragaman huruf, berkaitan dengan huruf kapital atau bukan [3]. Ini adalah contoh kalimat *case folding*, *input* "Bukalapak Bikin Layanan Pinjaman Tanpa Syarat untuk UKM", menjadi *output* "bukalapak bikin layanan pinjaman tanpa syarat untuk ukm".

5. *Tokenization*

Suatu proses yang dilakukan untuk memotong atau memecah kalimat menjadi beberapa bagian atau kata-kata. Hasil deduksi ini disebut token. Dalam beberapa kasus, proses *tokenization* juga dilakukan dengan menghilangkan tanda baca yang tidak diperlukan [8]. Ada beberapa model *tokenization* yang dapat digunakan, yaitu *unigram, bigram, trigram, dan ngram.* Ini adalah contoh kalimat *tokenization*, *input* "bukalapak bikin layanan pinjaman tanpa syarat untuk ukm", keluaran "bukalapak, bikin, layanan, pinjaman, tanpa, persyaratan, untuk, ukm".

6. *Filtering*

Filtering adalah tahap menghilangkan kata-kata yang muncul dalam jumlah besar tetapi dianggap tidak memiliki makna (*stopwords*). Pada dasarnya, daftar *stopwords* adalah sekumpulan kata yang banyak digunakan dalam berbagai bahasa. Alasan menghapus kata yang terkait dengan penambangan teks adalah karena penggunaannya yang terlalu umum, sehingga pengguna dapat fokus pada kata-kata lain yang jauh lebih penting [3]. Ini adalah contoh kalimat *stopwords*, *input* "bukalapak bikin layanan pinjaman tanpa syarat untuk ukm", *output* "bukalapak bikin layanan pinjaman tanpa syarat ukm". Ini adalah salah contoh kata dari *Stopwords*:

ada	di	kalau	pada	yaitu
aku	dia	kami	saja	bila
bapak	ini	lalu	tentu	hari
berbagai	itu	lewat	untuk	masa
Cara	jadi	meski	yang	tapi
Cuma	juga	oleh	wah	hal

Tabel 3. Contoh Kata *Stopword*

7. Stemming

Stemming adalah tahap untuk menjadikan kata suffix menjadi kata dasar sesuai dengan aturan Indonesia yang benar [8]. Ini adalah contoh dari kalimat, input "bukalapak bikin layanan pinjaman tanpa syarat untuk ukm", dan output "bukalapak bikin layan pinjam syarat ukm". Berikut ini adalah salah satu contoh kata dari stemming:

Sebelum	**Sesudah**
adanya	ada
akhiri	akhir
sebelum	belum
diberikan	beri
secukupnya	cukup
dipergunakan	guna

Tabel 4. Contoh Kata *Stemming*

8. *Weighting Word*

Weighting Word adalah mekanisme untuk memberi skor pada frekuensi kemunculan kata dalam dokumen teks. Salah satu metode populer untuk menimbang kata adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). *Term Frequency-Inverse Document Frequency* adalah metode pembobotan yang menggabungkan dua konsep, yaitu *Term Frequency* dan *Document Frequency*. Istilah Frekuensi adalah konsep pembobotan dengan mencari seberapa sering (frekuensi) penampilan suatu istilah dalam satu dokumen. Karena setiap dokumen memiliki panjang yang berbeda, dapat terjadi bahwa sebuah kata muncul lebih banyak di dokumen panjang dibandingkan dengan dokumen pendek. Dengan demikian, frekuensi istilah sering dibagi dengan panjang dokumen (total kata dalam dokumen) [7].

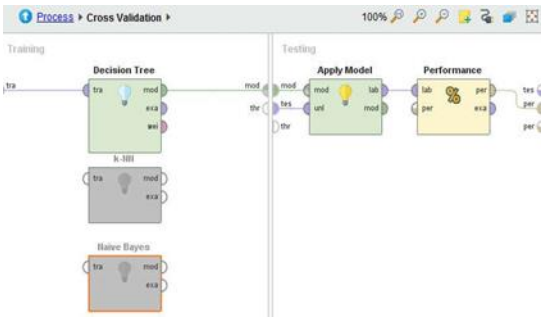
Sedangkan Frekuensi Dokumen adalah jumlah dokumen tempat *term* muncul. Semakin kecil frekuensi kemunculannya, semakin kecil nilainya. Saat menghitung frekuensi istilah, semua kata di dalamnya dianggap penting. Namun, ada kata-kata yang sebenarnya kurang penting dan tidak perlu diperhitungkan seperti "di-", "kem", "dan", dll. Oleh karena itu, kata-kata yang kurang penting ini perlu dikurangi dan menambah dengan kata-kata penting lainnya. Ini adalah ide dasar mengapa *stopword* dibutuhkan. Dengan demikian diperlukan TF-IDF untuk menghitung, skor dapat diperoleh dengan menggunakan Persamaan [3].

$$tf - idf_{td} = tf_{td}\, x\, idf_t$$

Frekuensi istilah (tf) adalah frekuensi kemunculan istilah (t) dalam dokumen (d) [8].

C. Klasifikasi Analisis Sentimen

Setelah pra-pemrosesan data, langkah selanjutnya adalah klasifikasi analisis sentimen. Tahap ini adalah tahap untuk memberikan pelatihan dan mengimplementasikan berbagai algoritma penambangan data [3]. Gambar 5 menunjukkan konten operator "*Cross Validation*" dalam aplikasi *Rapidminer*. Dalam hal ini, menggunakan tiga operator klasifikasi berbeda untuk perbandingannya, yaitu operator klasifikasi *Decision Tree*, operator klasifikasi K-NN dan operator klasifikasi *Naïve Bayes*.



Gambar 5. Cross Validation Operator

D. Evaluasi Analisis Sentimen

Setelah proses klasifikasi analisis sentimen selesai, diperlukan satu langkah lagi untuk menentukan kualitas proses yang telah dilakukan, yaitu mengevaluasi hasilnya. Pada tahap ini, kinerja perhitungan yang telah dilakukan akan diuji dengan parameter *accuracy, precision,* dan *recall.*

Akurasi (A) adalah jumlah dokumen yang diklasifikasikan dengan benar, baik Benar Positif maupun Benar Negatif. Menghitung nilai akurasi dapat menggunakan persamaan:

$$A = \frac{(TP+TN)}{(TP+FP+TN+FN)} x100\%$$

Presisi (P) adalah seberapa banyak hasil pemrosesan relevan dengan informasi yang ingin dicari. Dengan kata lain, presisi adalah klasifikasi *True Positive* dan semua data diprediksi sebagai kelas positif. Menghitung nilai presisi dapat menggunakan persamaan:

$$P = \frac{TP}{(TP+FP)} x100\%$$

Recall (R) adalah berapa banyak dokumen yang relevan dalam koleksi dihasilkan oleh sistem. Dengan kata lain, *recall* adalah jumlah dokumen yang memiliki klasifikasi Benar Positif dari semua dokumen yang benar-benar positif (termasuk *False Negative*). Menghitung nilai *recall* dapat menggunakan persamaan:

$$P = \frac{TP}{(TP+FN)} x100\%$$

Variabel seperti TP, TN, FP, dan FN berasal dari *confusion matrix*. TN singkatan dari *True Negative*, data negatif diklasifikasikan sebagai negatif. TP merupakan singkatan dari *True Positive*, data positif diklasifikasikan sebagai positif. FN adalah singkatan dari *False Negative*, data positif diklasifikasikan sebagai negatif. FP singkatan *False Positive* [3], data negatif diklasifikasikan sebagai positif. Untuk penjelasan yang lebih detail:

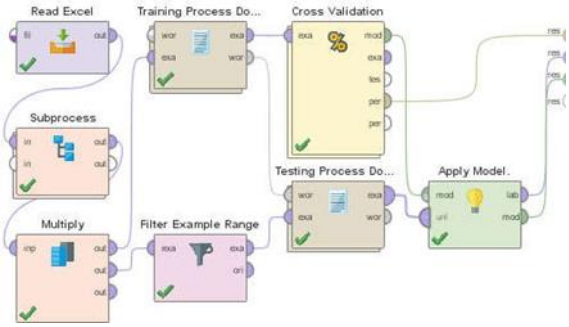
	Prediction Yes	**Prediction No**
True Yes	*TP*	*FN*
True No	*FP*	*TN*

Tabel 5. *Confusion Matrix*

Setelah data dikumpulkan, data akan dibagi menjadi data *training* dan data *testing*. Pembagian data akan dilakukan dengan menggunakan metode validasi *N-fold cross* untuk menghilangkan bias kata. *N-fold cross* validasi membagi dokumen menjadi n bagian. Dalam serangkaian eksperimen akan dilakukan n potongan eksperimen klasifikasi dokumen dengan setiap eksperimen menggunakan satu bagian sebagai pengujian data, (n-1)/2 bagian sebagai dokumen berlabel, dan (n-1)/2 bagian lain sebagai dokumen tanpa label yang akan ditukar setiap eksperimen sebanyak n kali. Kumpulan dokumen yang dimiliki pertama secara acak diurutkan sebelum dimasukkan ke dalam lipatan. Hal ini dilakukan untuk menghindari pengelompokan dokumen dari satu kategori tertentu di lipatan [8].

3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil hasil eksperimen dan menganalisis kinerjanya.



Gambar 6. Proses Utama pada *Rapidminer*

Gambar 6 menunjukkan proses utama dalam aplikasi *Rapidminer*. Konten "*Read Excel*" digunakan untuk membaca data dalam file *Excel*. Konten "*Subprocess*" dan "*Process Document*" digunakan untuk pra-pemrosesan. Konten "*Cross Validation*" digunakan untuk klasifikasi dan evaluasi analisis sentimen dengan eksperimen yang dilakukan sepuluh kali (10-lipat *Cross Validation*) [8].

Sentiment	prediction(Sentiment)	co...	...	text
Negative	Negative	1	0	bukalapak bukalapak tolong tanggapi kecewa dengn bukalapak
Negative	Negative	1	0	min sampe tindakan apa admin tokped penjual mengulur nih
Positive	Positive	0.118	0.882	btw pas ngetik keywords sedotan tokopedia hasil sedotan stainless senang yg peduli senang
Positive	Positive	0.118	0.882	bukalapak bikin layanan pinjaman syarat ukm

Gambar 7. Hasil Prediksi *Rapidminer*

Gambar 7 menunjukkan hasil prediksi dalam aplikasi *Rapidminer*. Melihat sentimen dan juga label prediksi yang sama dengan label sentimen.

Berikut ini gambaran hasil dari *confusion matrix* dari masing-masing algoritma pada *Rapidminer*:

Metode	***TP***	***FP***	***TN***	***FN***
Decision Tree	42	12	38	8
K-NN	35	7	43	15
Naïve Bayes	32	5	45	18

Tabel 6. *Confusion Matrix* Setiap Algoritma

Dari *confusion matrix* pada tabel 6 nilai rata-rata *accuracy*, *precicion* dan *recall* ditampilkan dengan perhitungan menggunakan rumus.

Metode	**Accuracy**	**Precision**	**Recall**
Decision Tree	82%	80%	86%
K-NN	80%	85%	72%
Naïve Bayes	79%	88%	66%

Tabel 7. Nilai *Accuracy*, *Precision*, *Recall* Menggunakan Formula

Ada perbedaan dalam hasil nilai rata-rata presisi menggunakan aplikasi *Rapidminer* seperti yang ditunjukkan pada tabel di bawah ini:

Metode	**Accuracy**	**Precision**	**Recall**
Decision Tree	82 %	81.95 %	86 %
K-NN	80 %	87.68 %	72 %
Naïve Bayes	79 %	89.50 %	66 %

Tabel 8. Nilai *Accuracy*, *Precision*, *Recall* Menggunakan *Rapidminer*

Hasil ini menunjukkan *accuracy* dari *Decision Tree*, KNN, dan *Naïve Bayes* sebesar 82%, 80%, dan 79%. Hasil untuk *precision* dari *Decision Tree*, K-NN, dan *Naïve Bayes* sebesar 81,95%, 87,68%, dan 89,50%. Sementara hasil untuk *Recall* dari *Decision Tree*, K-NN, dan *Naïve Bayes* adalah 86%, 72%, dan 66%. Sehingga dapat dilihat bahwa *classifier Naïve Bayes* adalah *classifier* terbaik untuk digunakan dengan dataset media sosial karena memberikan prediksi yang lebih akurat dan tepat. Ada perbedaan dalam hasil penelitian sebelumnya, menunjukkan akurasi dari K-NN, *Naïve Bayes*, dan *Decision Tree classifier* sebesar 84,66%, 50,72%, dan 64,42%. Perbedaannya terjadi karena karakteristik dari berbagai dataset dan proses.

4. PENUTUP

KESIMPULAN

Dalam studi ini, upaya dilakukan untuk mengklasifikasikan sentimen analisis dari tweet di Twitter tentang pendapat pelanggan tentang beberapa situs Marketplace di Indonesia. Untuk merangkum pandangan pelanggan ini teknik penambangan teks digunakan, dan penggalian data menggunakan tiga algoritma yang berbeda yaitu Decision Tree, K-NN, dan NaïveBayes. Tiga algoritma memprediksi label pada dataset. Hasilnya menunjukkan accuracy dari Decision Tree, K-NN, dan Naïve Bayes yaitu 82%, 80%, dan 79%. Hasil untuk precision dari Decision Tree, K-NN, dan Naïve Bayes sebesar 81,95%, 87,68%, dan 89,50%. Hasilnya juga menunjukkan bahwa Recall dari Decision Tree, K-NN, dan Naïve Bayes adalah 86%, 72%, dan 66%. Sehingga dapat disimpulkan bahwa algoritma Decision Tree adalah classifier terbaik untuk digunakan dalam perhitungan yang menggunakan dataset media sosial karena memberikan prediksi yang lebih akurat dan tepat.

SARAN

Di masa mendatang, kita harus menggunakan dataset yang lebih besar dan lebih kompleks, hal tersebut dapat diwujjudkan dengan meningkatkan jumlah label serta jangkauan yang lebih luas dari hanya sekedar *Marketplace.* Akan lebih baik lagi jika kita dapat menggunakan standar diluar Indonesia.

DAFTAR PUSTAKA

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," pp. 30–38, 2011.

[2] S. Rodiyansyah, "Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan," *Infotech J.*, vol. 1, no. 2, p. 236599, 2015.

[3] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree-Knn-Naïve Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.

[4] S. Al-Osaimi and M. Badruddin, "Sentiment Analysis Challenges of Informal Arabic Language," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 2, pp. 278–284, 2017, doi: 10.14569/ijacsa.2017.080237.

[5] Ramadina, "Penerapan Fungsi Data Mining Klasifikasi Untuk Prediksi Masa Studi Mahasiswa Tepat Waktu Pada Sistem Informasi Akademik Perguruan Tinggi," *JUPITER (Jurnal Penelit. Ilmu dan Teknol. Komputer)*, vol. 7, no. 1, pp. 39–50, 2015.

[6] F. Laeeq, M. T. Nafis, and M. R. Beg, "Sentimental Classification of Social Media using Data Mining," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 546–549, 2017.

[7] M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 57–64, 2016.

[8] R. Rasenda, H. Lubis, and R. Ridwan, "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 369, 2020, doi: 10.30865/mib.v4i2.2051.