

Sentimen Analisis *Social CRM* pada Media Sosial *Instagram* Menggunakan *Machine Learning* untuk Mengukur Retensi Pelanggan

Qhairani Frilla F. Safiesza*, M Afdal, Rice Novita, Mustakim

Fakultas Sains & Teknologi, Sistem Informasi, UIN Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: 12050322811@students.uin-suska.ac.id, m.afdal@uin-suska.ac.id, ricenovita@uin-suska.ac.id, mustakim@uin-suska.ac.id

Email Penulis Korespondensi: 12050322811@student.uin-suska.ac.id

Submitted: 03/06/2024; Accepted: 29/06/2024; Published: 07/09/2024

Abstrak—Guna menciptakan dan mempertahankan daya saing yang unggul dalam perekonomian berbasis pengetahuan maka sebuah bisnis harus mampu memanfaatkan data serta mengelola hubungan terhadap pelanggan lewat penerapan Customer Relationship Management khususnya Social CRM. Social CRM menjadi sebuah pembaharuan dalam strategi bisnis yang diciptakan untuk melibatkan pelanggan dalam sebuah percakapan kolaboratif dan menciptakan nilai yang saling menguntungkan dalam suatu lingkungan bisnis yang terpercaya dan transparan. Melihat perkembangan ini sebagai salah satu perusahaan kuliner di sektor Oleh-oleh yang sukses di pekanbaru, perusahaan harus mampu untuk memproses semua informasi yang didapatkan. Saat ini perusahaan belum pernah menganalisis komentar di media sosial terkhususnya akun instagram. Komentar komentar ini berguna untuk bahan evaluasi serta dapat menjadi parameter kepuasan pelanggan serta melihat potensi retensi pelanggan. Untuk menilai komentar komentar positif dan negatif pada akun instagram tersebut maka dapat dilakukan menggunakan sentimen analisis dengan machine learning yakni 3 algoritma klasifikasi yaitu Naive Bayes Classifier (NBC), Support Vector Machine(SVM) dan Random Forest (RF). Hasil sentiment menunjukkan algoritma SVM dan NBC memperoleh akurasi terbaik sebesar 74.26% dibandingkan RF serta hasil analisis social CRM menunjukkan pelanggan lebih puas terhadap perusahaan dari segi produk, pelayanan, dan aksi yang dilakukan oleh perusahaan sehingga perusahaan dianggap mampu mempertahankan pelanggannya.

Kata Kunci: Machine Learning, Naive Bayes Classifier, Random Forest, Retensi Pelanggan Sentimen Analisis, Social CRM, Support Vector Machine

Abstract—To create and maintain a superior competitive advantage in a knowledge-based economy, businesses must be able to utilize data and manage customer relationships through the implementation of Customer Relationship Management (CRM), particularly Social CRM. Social CRM is a renewal of business strategy that is created to engage customers in a collaborative conversation and create mutually beneficial value in a trusted and transparent business environment. Seeing this development as one of the successful culinary companies in the Souvenir sector in Pekanbaru, the company must be able to process all the information obtained. Currently, the company has never analyzed comments on social media, especially the Instagram account. These comments are useful for evaluation material and can be a parameter of customer satisfaction and to see the potential for customer retention. To assess positive and negative comments on the Instagram account, sentiment analysis can be carried out using machine learning, namely 3 classification algorithms, namely Naive Bayes Classifier (NBC), Support Vector Machine (SVM) and Random Forest (RF). The sentiment results show that the SVM and NBC algorithms obtain the best accuracy of 74.26% compared to RF, and the results of the social CRM analysis show that customers are more satisfied with the company in terms of products, services, and actions taken by the company, so that the company is considered capable of retaining its customers.

Keywords: Machine Learning, Naive Bayes Classifier, Random Forest, Customer Retention, Analysis Sentiment, Social CRM, Support Vector Machine

1. PENDAHULUAN

Dalam perekonomian berbasis pengetahuan yakni di era bisnis yang canggih dan berkembang saat ini, salah satu faktor keberhasilan daya saing yang paling penting di antara perusahaan adalah manajemen pengetahuan pelanggan[1][2]. Selain itu perusahaan pasti menghadapi tantangan yang luar biasa seperti tingginya persaingan kompetitif antar perusahaan, perubahan trend konsumsi yang sulit untuk diprediksi dan banyak lagi tuntutan pelanggan yang canggih[3][4] termasuk di sektor kuliner terkhususnya oleh-oleh. Untuk menciptakan dan mempertahankan daya saing yang unggul ini perusahaan perlu memberikan layanan yang unggul terfokus di bidang pemasaran dengan menerapkan Customer Relationship Management khususnya Social CRM[5][6].

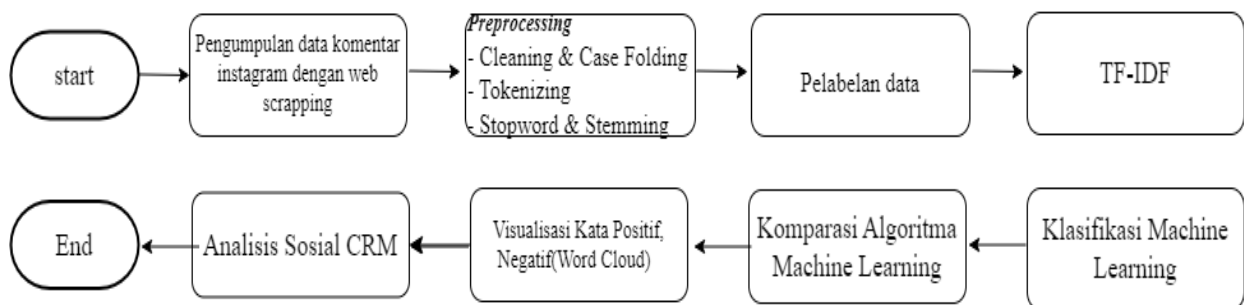
Social CRM didefinisikan sebagai "aktivitas mengintegrasikan segala sesuatu yang berhubungan dengan pelanggan termasuk proses, sistem, dan teknologi". hal ini berbanding lurus dengan munculnya aplikasi media sosial yang melibatkan pelanggan dalam percakapan kolaboratif yang dapat meningkatkan hubungan bisnis terhadap pelanggan[7][8][9]. Media sosial memaparkan banyak sumber informasi bagi sebuah bisnis seperti suka, komentar, dan tagar yang mampu menjadi informasi berharga bagi perusahaan untuk meningkatkan jumlah pelanggan[7][10]. Instagram kini menjadi salah satu media sosial yang populer saat ini dengan total pengguna 116,16 juta pengguna di Indonesia hingga Agustus 2023[10][11]. Penelitian yang dilakukan oleh Jalal, Bahari, Tarofder (2021) menunjukkan bahwa konsep Social CRM ini memiliki dampak yang luar biasa pada organisasi layanan kesehatan sehingga penyedia layanan kesehatan mampu mengembangkan teknik pemasaran yang efektif dan memajukan manfaat yang dirasakan masyarakat lewat Sosial CRM[11] serta Penelitian Harrigan (2020) juga membahas bagaimana perbedaan antara Social CRM dan Tradisional CRM dan menyatakan media sosial ialah sumber daya terbaik yang digunakan oleh organisasi untuk mengkonsepkan informasi dan pemasaran pada tingkat yang lebih strategis serta taktis di dalam Social CRM itu sendiri[2].

Sebagai salah satu perusahaan kuliner di sektor Oleh-oleh yang sukses di pekanbaru, pelaku bisnis harus mampu mengolah seluruh informasi yang diterimanya, mengenali apa yang telah terjadi dan memprediksi apa yang akan terjadi di masa depan[7].Hingga saat ini, perusahaan belum menganalisis komentar di akun instagram. Komentar tersebut dapat digunakan untuk mengevaluasi perusahaan agar mampu memberikan pelayanan yang terbaik kepada seluruh pelanggannya[12].Komentar ini juga dapat digunakan sebagai penilaian kepuasan pelanggan terhadap produk[13]Untuk menilai dan mengevaluasi komentar positif dan negatif tersebut dapat di uji coba menggunakan analisis sentimen[13][14]dengan menggunakan teknik scraping[15]Sebagaimana penelitian S. Lamrhari (2020) yang membahas pentingnya manajemen hubungan pelanggan bagi perusahaan serta mengelompokkan pelanggan ke dalam kelas tertentu menggunakan Machine Learning yakni algoritma Random Forest serta dibandingkan dengan algoritma Support Vector Machine dan K-Nearest Neighbors. Perbedaan pada penelitian sebelumnya yakni terletak pada tujuan dan metode yang digunakan yakni untuk mengukur retensi pelanggan menggunakan 3 algoritma Machine Learning yaitu Naive Bayes, Support Vector Machine dan Random Forest[15]

Analisis sentimen (*sentiment analysis*) adalah sebuah studi dalam bidang komputasi yang menggunakan opini, pandangan dan perasaan yang direpresentasikan ke dalam bentuk tekstual[6][16].Konsep dasar dari analisis sentimen adalah proses pengklasifikasian data menjadi opini positif dan opini negative. Penelitian yang dilakukan oleh Ratino, Noor dkk (2020) membahas penerapan algoritma Support Vector Machine dan Naive Bayes Classifier dalam mengetahui sentimen terhadap informasi Covid-19 pada komentar sosial media instagram dengan hasil akurasi 80,23% diperoleh dari SVM dan 78,02% dari Naive Bayes Classifier[17].Selanjutnya, penelitian yang dilakukan oleh Wahyu Surya Gemilang (2024) membahas bagaimana memahami pandangan masyarakat dengan menganalisis sentimen pengguna instagram terkait calon presiden 2024 dengan menggunakan SVM dengan hasil berupa wawasan tentang sentimen masyarakat yang diharapkan dapat memberi manfaat bagi pemangku kepentingan dan masyarakat umum[16]. Penelitian oleh Mega dkk(2020) juga membahas penerapan algoritma Naive Bayes Classifier dalam menganalisis komentar instagram gubernur jawa barat dan memperoleh akurasi sebesar 92%[18].Teknik klasifikasi yang akan digunakan pada penelitian ini adalah algoritma Naive Bayes (NB), Support Vector Machine (SVM), dan Random Forest[19]. Ketiga algoritma klasifikasi ini dinilai memiliki keunggulannya masing-masing sehingga akan menghasilkan akurasi terhadap analisis sentimen komentar pada akun instagram perusahaan[20][21]

2. METODOLOGI PENELITIAN

Adapun metodologi penelitian dalam penelitian ini secara garis besar terdiri atas 7 tahapan yang dilakukan. Berikut ini adalah tahap-tahap dalam penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Dalam penelitian ini dilakukan sentiment analysis *Social CRM* pada salah satu perusahaan yang bergerak di bidang oleh oleh dan kuliner di Pekanbaru. Adapun data yang diambil berupa data komentar instagram dengan text mining untuk mendapatkan data komentar selama 1 tahun terakhir (Januari-Desember 2023). Kemudian data yang telah dikumpulkan tersebut selanjutnya masuk ke tahap pengolahan data yakni *text processing* mulai dari *case folding* untuk mengubah teks menjadi bentuk standar yakni menjadi huruf kecil, *tokenizing* untuk pemotongan kalimat menjadi sebuah kata, *stopword* untuk menghilangkan kata-kata umum yang sering muncul seperti “yang”, “dan”, “di”, lalu *stemming* yakni untuk menyelaraskan kata untuk mereduksi daftar kata dalam data latih. Kemudian dilakukan ekstraksi fitur menggunakan TF-IDF dan setelahnya dilakukan pelabelan yakni komentar positif, komentar negatif dan komentar netral.

Tahap selanjutnya yaitu pemodelan menggunakan machine learning yakni algoritma Naive Bayes Classifier, Support Vector Machine dan Random Forest yang kemudian akan diuji performanya menggunakan *confusion matrix*. Kemudian akan diperoleh akurasi dari masing-masing algoritma dan selanjutnya masuk ke tahap visualisasi dengan menggunakan *word cloud*. Kemudian analisis *social CRM* tersebut dapat diimplementasikan terhadap perusahaan agar dapat dijadikan pembelajaran serta bahan evaluasi untuk strategi kedepannya.

2.1 Analisis Sentimen

Analisis sentimen adalah proses mengklasifikasi dokumen teks yang berupa opini berdasarkan sentimen untuk menentukan apakah suatu tanggapan bersifat positif, netral atau negatif[22].Tujuan dari analisis sentimen untuk



menentukan opini dari seseorang dengan memperhatikan topik tertentu dan melakukan ekstraksi atribut pada suatu dokumen atau teks yang berisi komentar untuk mengetahui ekspresi yang ada di dalamnya sehingga komentar tersebut dapat dikategorikan positif atau negatif[23]. Langkah yang dilakukan dalam analisis sentimen adalah mendefinisikan domain dataset, preprocessing, featured selection, pelabelan, klasifikasi dan evaluasi[24]. Konsep dasar dari analisis sentimen adalah proses pengklasifikasian data menjadi opini positif dan opini negative. Beberapa teknik klasifikasi yang dapat digunakan ialah Naive Bayes(NB), Support Vector Machine(SVM), dan Random Forest.

2.2 Algoritma Naive Bayes Classifier (NBC)

Algoritma Naive Bayes Classifier merupakan sebuah teknik dalam klasifikasi probabilistik yang sederhana dimana dalam teknik ini dilakukan perhitungan probabilitas dengan cara melakukan penjumlahan terhadap frekuensi dan kombinasi nilai dari sebuah dataset[24]. Menurut Rossi, A.dkk Naive Bayes Classifier adalah konsep probabilitas penentu kelompok kelas dokumen[18]. Algoritma klasifikasi ini dapat mengolah data dalam jumlah besar dengan hasil akurasi yang tinggi. Adapun keuntungan dengan menggunakan metode Naive Bayes Classifier karena metode Naive Bayes Classifier memerlukan data latih yang sedikit dalam proses klasifikasi teks[24]. Berikut Metode Naive Bayes dapat dituliskan pada persamaan berikut ini

$$\frac{P(h|e)=P(e|h).P(h)}{P(e)} \tag{1}$$

$P(h|e)$ ialah Probabilitas akhir dimana hipotesis h terjadi jika diberikan bukti e . Kemudian $P(e|h)$ berarti Probabilitas bukti e terjadi akan mempengaruhi hipotesis h dan $P(h)$ adalah Probabilitas awal h tanpa memandang bukti serta $P(e)$ yaitu Probabilitas awal e terjadi tanpa memandang bukti lain.

2.3 Algoritma Support Vector Machine (SVM)

Support Vector Machine adalah sistem pembelajaran mesin yang menggunakan fungsi-fungsi linear dalam fitur dengan dimensi yang tinggi. SVM dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi[25]. Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan pada SVM sangat bergantung pada fungsi kernel dan parameter yang digunakan. SVM sangat cepat dan efektif untuk masalah klasifikasi teks, dalam istilah geometris sebuah klasifikasi biner dapat dilihat sebagai *Hyperplane* dalam ruang fitur yang memisahkan titik-titik yang mewakili contoh positif dan kategori yang mewakili keadaan negative[26] Prinsip kerja SVM itu sendiri yakni dengan menemukan hyperplane yang bisa melakukan klasifikasi data menjadi 2 kelas. Adapun rumus perhitungannya adalah sebagai berikut[27] :

$$\{(X_i, Y_i)\}_{i=1}^N \tag{2}$$

Maksimalkan fungsi :

$$Ld = \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \text{ syarat : } 0 \leq \alpha_i \leq C \text{ dan } \sum_{i=1}^N \alpha_i y_i = 0 \tag{3}$$

Menghitung nilai w dan b :

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \tag{4}$$

Fungsi keputusan klasifikasi sign (f(x)) :

$$w(x) = w \cdot x + b \text{ atau } f(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b \tag{5}$$

N merupakan jumlah data sedangkan n adalah banyak fitur atau dimensi pada data. Adapun C adalah nilai konstanta, m yaitu jumlah support vector atau titik data yang memiliki $\alpha_i > 0$ dan $K(x, x_i)$ yaitu fungsi kernel. Adapun Ld adalah dualitas lagrange multiplier dan α_i berupa nilai bobot setiap titik data.

2.4 Algoritma Random Forest

Random Forest adalah salah satu algoritma klasifikasi yang dibangun dari beberapa pohon keputusan dan menggabungkannya untuk mendapatkan prediksi yang lebih stabil dan akurat yang diusulkan oleh Breiman[20]. Random forest bertujuan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan mengambil atribut dan data secara acak sesuai ketentuan. Adapun kelebihan dari Random Forest ialah perlindungan terhadap overfitting yang membuat model mampu memberikan hasil dengan kinerja yang tinggi[28]. Root node adalah simpul yang terletak di atas (akar dari pohon keputusan. Sedangkan internal node merupakan simpul percabangan, node ini mempunyai satu input dan minimal dua output. Leaf node/terminal node yaitu simpul terakhir yang tidak memiliki output dan hanya memiliki satu input. Perhitungan nilai entropy menggunakan rumus pada persamaan 1, dan nilai information gain pada rumus persamaan 2[29]

$$Entropy(Y) = \sum_i p(c|Y) \log_2 p(c|Y) \tag{6}$$

Y merupakan himpunan kasus dan $p(c|Y)$ adalah proporsi nilai Y pada kelas c .

Information Gain (Y, a)



$$Entropy(Y) = \sum_{v \in Values(a)} \frac{|Y_v|}{|Y_a|} Entropy(Y_v) \tag{7}$$

Values(a) adalah semua nilai yang mungkin pada himpunan kasus a. Y_v merupakan subkelas dari Y dengan kelas v yang berkaitan dengan kelas a. Y_a merupakan semua nilai yang sama dengan a.

HASIL DAN PEMBAHASAN

Pada data yang digunakan dalam penelitian ini diambil dari komentar-komentar postingan instagram perusahaan. Pengambilan data menggunakan teknik scraping pada instagra. Proses scrapping menghasilkan 1945 data yang merupakan komentar dari postingan instagram perusahaan selama 1 tahun terakhir (Januari-Desember 2023). Seluruh data kemudian dilakukan preprocessing yang terdiri dari beberapa proses yaitu cleaning, case folding, tokenizing, stopword & stemming. Pada proses cleaning semua emoticon, mention, hashtag pada data komentar akan dihapus. selanjutnya pada proses case folding semua teks akan diubah menjadi huruf kecil. Kemudian pada proses tokenizing kalimat akan dipecah menjadi kata per kata. Selanjutnya pada proses stopword akan dihapus kata-kata yang bukan merupakan kata kunci pada komentar dan terakhir pada tahapan stemming akan diselaraskan kalimat-kalimat yang ada untuk mengurangi daftar kata pada data training. Hasil dari tahapan preprocessing dapat dilihat pada Tabel 1.

Tabel 1. Hasil Preprocessing Data

Tahapan	Hasil
Data Awal	Keren moga viera buka cabang daerah di riau biar bantu ukm kenal khas riau dan wisatawan riau. tunggu bukak cabang di bengkal y bun sayo nak lamar karyawan ni. sukses viera varian nya enak suka belanja di viera ramah ramah. berkah selau yaa bun murah rezeki nya.
Cleaning & Casefolding	keren semoga viera buka cabang di daerah riau biar bantu ukm kenal khas riau wisatawan riau. tunggu bukak cabang bengkal bun sayo nak lamar karyawan ni. sukses viera varian enak suka belanja viera ramah ramah. berkah selalu yaa bun murah rezeki
Tokenizing	['keren','semoga','viera','buka','cabang','daerah','riau','biar','bantu','ukm','kenal','khas','riau','wisatawan','riau','tunggu','bukak','cabang','bengkal','bun','sayo','nak','lamar','karyawan','ni','sukses','viera','varian','enak','suka','belanja','viera','ramah','ramah','berkah','selalu','yaa','bun','murah','rezeki']
Stopword/Filtering	['keren','moga','viera','buka','cabang','daerah','riau','biar','bantu','ukm','kenal','khas','riau','wisatawan','riau','tunggu','bukak','cabang','bengkal','bun','sayo','nak','lamar','karyawan','ni','sukses','viera','varian','enak','suka','belanja','viera','ramah','ramah','berkah','selalu','yaa','bun','murah','rezeki']
Stemming	['keren','moga','viera','buka','cabang','daerah','riau','biar','bantu','ukm','kenal','khas','riau','wisatawan','riau','tunggu','bukak','cabang','bengkal','bun','sayo','nak','lamar','karyawan','ni','sukses','viera','varian','enak','suka','belanja','viera','ramah','ramah','berkah','selalu','yaa','bun','murah','rezeki']

3.1 Pelabelan Data

Pelabelan ini dilakukan dengan cara memberikan nilai positif, negatif dan juga netral. Pelabelan ini dilakukan guna untuk mencapai target yang diinginkan sebelum dilakukan analisis menggunakan TF-IDF dan klasifikasi *Naive Bayes Support Vector Machine* dan *Random Forest*. Pelabelan ini dilakukan oleh pakar ahli sastra bahasa Indonesia. Adapun potongan hasil pelabelan dapat dilihat pada Tabel 2.

Tabel 2. Pelabelan

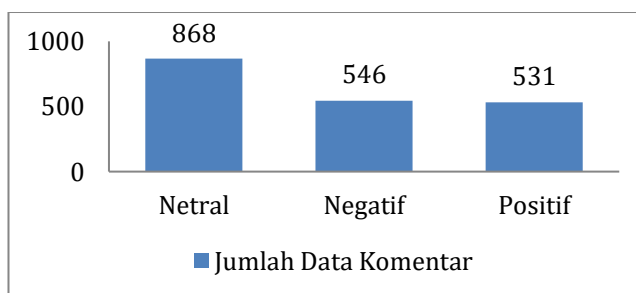
No	Komentar	Sentimen
1	keren moga viera buka cabang daerah riau biar bantu ukm kenal khas riau wisatawan riau. tunggu bukak cabang bengkal bun sayo nak lamar karyawan ni. sukses viera varian enak suka belanja viera ramah ramah. berkah selau yaa bun murah rezeki	Positif
2	salah cari lawan micue	Negatif
3	awaliyahh terimakasih kakak kakak the best	Netral
...
1945	coklat ganti silver queen ja bun produk asli indonesia	Positif

Setelah dilakukan pelabelan, data tersebut diklasifikasikan dan digunakan sebagai parameter dalam menganalisa *social CRM*. Adapun klasifikasi tersebut yakni komentar positif, negatif dan netral. Berikut potongan hasil komentar dapat dilihat pada Tabel 3.

Tabel 3. Pengklasifikasian Data Komentar

Username	Komentar	Sentimen
@cynthiaadewii	Yg paling bikin senang kalo belanja ke viera oleh-oleh adalah pelayanannya bagus & ramah2 karyawannya... tolong dipertahankan ya bunda karena ga semua kue enak dijualin ditoko oleh2 yg punya pelayanan ramah begini	Positif
@diabloseus	Kenapa di stop? Ada hubungannya kah?	Negatif
@aihhris	Mulai tgl berapa dan sampai tgl berapa bg	Netral

Komentar positif ini adalah komentar-komentar pelanggan yang berisi kepuasan pelanggan, pujian serta saran. Komentar netral ini diambil dari komentar-komentar pelanggan yang berisi komentar biasa berupa pertanyaan atau hal hal yang ingin didapatkan pelanggan berupa informasi lebih rinci terkait produk dan layanan sehingga dapat dilihat bahwasanya follower ingin menanyakan informasi yang lebih *detail*. Komentar negative ini diambil dari komentar-komentar pelanggan yang berisi kritikan, *complain*, dan ketidakpuasan terhadap produk atau layanan. Dari komentar negative tersebut dapat dilihat bahwa terdapat ketidakpuasan pelanggan. Setelah di klasifikasi, didapatkan keseluruhan data komentar. Rekapitulasi hasil komentar tersebut dapat dilihat pada gambar 2



Gambar 2. Rekapitulasi Data Komentar

Berdasarkan gambar 2 diatas, terdapat 531 komentar positif yang berisikan kepuasan, pujian, dan saran dari pelanggan. Untuk komentar netral terdapat 865 data yang berisikan komentar biasa, pertanyaan, dan meminta informasi detail mengenai produk atau layanan. Kemudian terdapat 549 komentar negatif yang berisikan kritikan, complaint, dan ketidakpuasan pelanggan.

3.2 Term Frequency (TF) - Inverse Document Frequency (IDF)

TF-IDF ini dilakukan untuk mendapatkan nilai dari bobot sebuah kata. Penulis melakukan tahap ini menggunakan module sklearn pada bahasa pemrograman Python sesuai dengan data training dan data set. Hasil tahapan TF-IDF dapat dilihat pada Tabel 4.

Tabel 4. Potongan Data Bobot TF-IDF

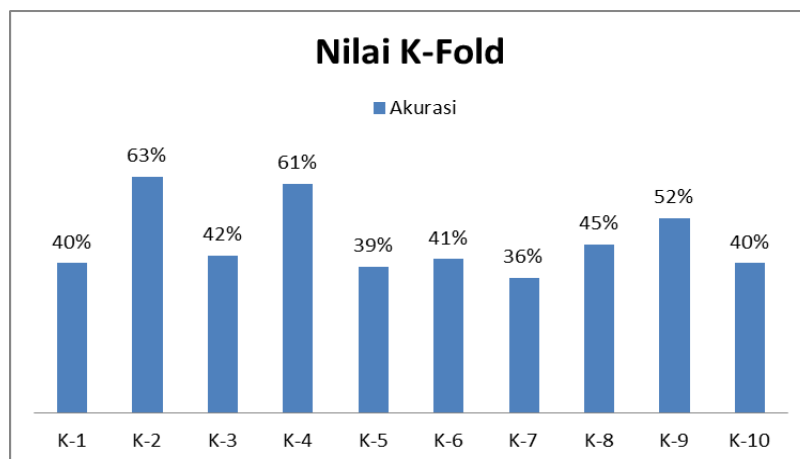
No	Abang	aamiin	aaaa	acara	abon	Adoyy
1	0.0	0.0	0.0	0.0	0.162	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.245	0.0	0.0	0.0	0.616
4	0.0	0.0	0.566	0.0	0.0	0.0
5	0.216	0.0	0.0	0.113	0.00	0.00
...
1945	0.0	0.0	0.00	0.00	0.00	0.00

3.3 Klasifikasi Machine Learning

Setelah melakukan *text preprocessing* dan *TF-IDF* maka tahapan selanjutnya adalah klasifikasi menggunakan *machine learning* yakni *naïve bayes*, *support vector machine* & *random forest*. Proses ini bertujuan untuk mengkategorikan data ke dalam sentiment positif dan sentiment negative.

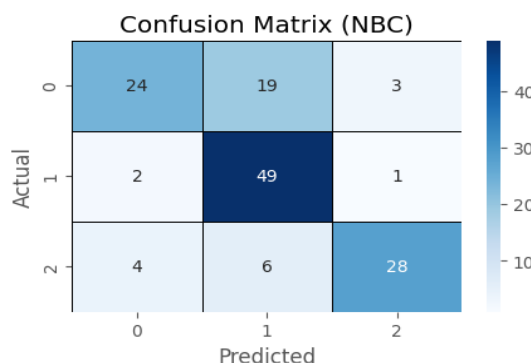
a. Klasifikasi Naïve Bayes

Pembagian data latih dan data uji dilakukan dengan teknik K-Fold Cross Validation. Adapun nilai K-Fold yang ditetapkan adalah 10 K-Fold. Data yang diuji sebanyak 1945 data. Pada hasil prediksi kelas data komentar menggunakan NBC akurasi tertinggi yaitu K-2 dengan nilai akurasi 63%. Hasil nilai K-Fold dapat dilihat pada gambar 3



Gambar 3. Nilai K-Fold

Untuk menguji kinerja *Naïve Bayes*, dilakukan pengujian menggunakan *confusion matrix*. Berikut hasil dari pengujian menggunakan *confusion matrix* dapat dilihat pada Gambar 4.



Gambar 4. Hasil Confusion Matrix Naive Bayes

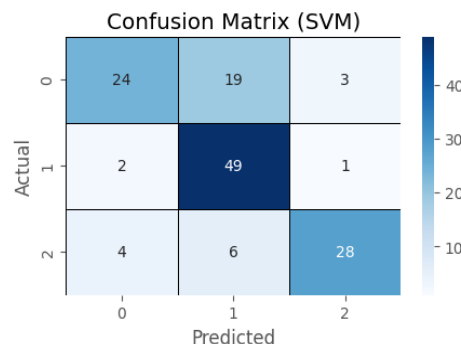
Hasil yang didapatkan dari pengujian ini mendapatkan nilai *accuracy* sebesar 60%. Selanjutnya dilakukan *Classification Report* dan diperoleh nilai *presisi*, *recall*, dan *f1-score* dapat dilihat pada tabel 5

Tabel 5. Hasil Classification Report

Hasil Pengujian <i>Naive Bayes</i>				
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.80	0.52	0.63	46
Netral	0.66	0.94	0.78	52
Positif	0.88	0.74	0.80	38

b. Klasifikasi Support Vector Machine

Implementasi algoritma *Support Vector Machine* dengan parameter *c*, kernel dan juga *gamma* memperoleh hasil berupa Parameter *C* yakni 100, Kernel *rbf* dan *Gamma* 0.1. Selanjutnya dilakukan *confusion matrix* untuk menguji kinerja Support Vector Machine. Berikut hasil pengujian menggunakan *confusion matrix* dapat dilihat pada Gambar 5



Gambar 5. Hasil Confusion Matrix Support Vector Machine

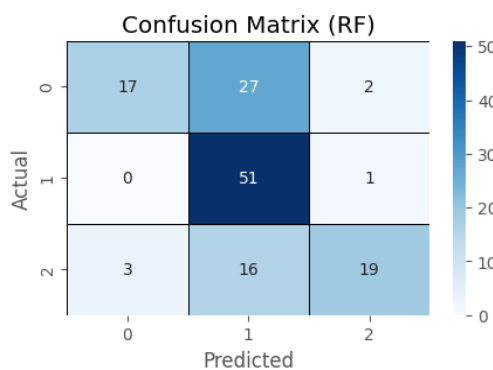
Hasil yang didapatkan dari pengujian ini mendapatkan nilai *accuracy* sebesar 76%. Selanjutnya dilakukan *Classification Report* dan diperoleh nilai *presisi*, *recall*, dan *f1-score* dapat dilihat pada tabel 6

Tabel 6. Hasil Classification Report

Hasil Pengujian Support Vector Machine				
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.80	0.52	0.63	46
Netral	0.66	0.94	0.78	52
Positif	0.88	0.74	0.80	38

c. Klasifikasi Random Forest

Untuk menguji kinerja *Random Forest* dilakukan pengujian menggunakan *confusion matrix*. Berikut hasil dari pengujian menggunakan *confusion matrix* dapat dilihat pada Gambar 6



Gambar 6. Hasil Confusion Matrix Random Forest

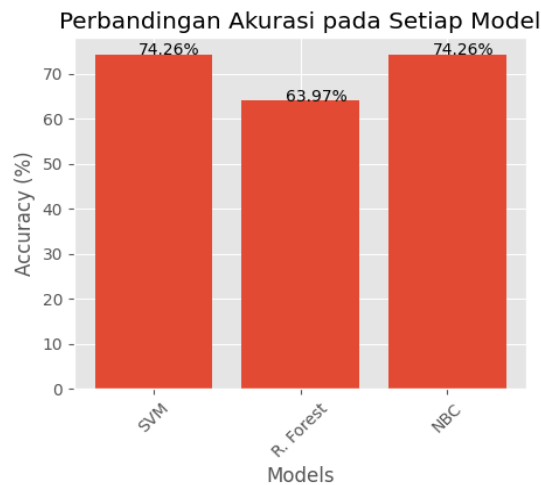
Hasil yang didapatkan dari pengujian ini mendapatkan nilai *accuracy* sebesar 60%. Selanjutnya dilakukan *Classification Report* dan diperoleh nilai *presisi*, *recall*, dan *f1-score* dapat dilihat pada tabel 7

Tabel 7. Hasil Classification Report

Hasil Pengujian Random Forest				
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.85	0.37	0.52	46
Netral	0.54	0.98	0.70	52
Positif	0.86	0.50	0.63	38

3.4 Perbandingan Model Klasifikasi

Setelah melewati proses validasi dengan ketiga algoritma yang ada yakni *Naïve Bayes Classifier*, *Support Vector Machine* dan *Random Forest* maka proses akurasi akan menghasilkan tingkat *Accuracy*, *recall*, *precision*, *f-1 score* dan *Support*. Hasil akurasi menunjukkan model algoritma SVM samal unggulnya dengan NBC dan lebih baik dibandingkan Random Forest yakni SVM dan NBC dengan akurasi 74.26%. kemudian Random Forest memiliki nilai akurasi yakni sebesar 63.97% Hasil perbandingan rata-rata setelah diproyeksikan dalam bentuk grafik dapat dilihat pada Gambar 7



Gambar 7. Perbandingan Akurasi Setiap Model

Berdasarkan perbandingan rata-rata masing-masing algoritma, SVM dan NBC lebih unggul ketimbang Random Forest, artinya SVM dan NBC memiliki performa terbaik.

3.5 Visualisasi Word Cloud

Tujuan visualisasi data adalah untuk mengekstrak informasi dalam bentuk topik yang paling sering didiskusikan pelanggan, sehingga memungkinkan perusahaan untuk mengekstrak informasi yang berpotensi penting dari sejumlah besar ulasan. Hasil analisis klasifikasi divisualisasikan menggunakan *Word Cloud*. *Word Cloud* adalah representasi data yang memperlihatkan kumpulan kata-kata penting yang sering muncul dalam data. Berikut hasil visualisasi kata pada data komentar instagram perusahaan yang menampilkan 100 kata secara keseluruhan dapat dilihat pada Gambar 8,9 dan 10



Gambar 8. Wordcloud kata keseluruhan



Gambar 9. Word Cloud kata positif



Username	Komentar	Jenis Komentar
@jumihandayani	Kue talamnya ada di balik logo Viera oleh-oleh bagian huruf C min. Ikutan kuisnya ya bestie @nniaaamell_@hesty_makeup_natural @lily. ummu.hanna	Give Away
@fia.fitria26	Haha....selalu ngakak lihat konten2nya. kerenlah pokoknya selalu menghibur	Soft Selling
@adelinazahida	Hanya khusus 3 produk ini ya min?	Diskon

Dari komentar-komentar tersebut dapat diindikasikan bahwa strategi yang perusahaan implementasikan berupa *Giveaway*, *Soft Selling* dan Diskon sangat menarik hati pelanggan dan membuat pelanggan bertahan dan berlangganan di perusahaan. Strategi tersebut membuat antusias pelanggan menjadi tinggi dan berdampak positif terhadap perusahaan.

3. KESIMPULAN

Dari penelitian yang dilakukan, dapat disimpulkan bahwa Hasil performa menggunakan 3 algoritma yakni *Naive Bayes Classifier*, *Support Vector Machine*, dan *Random Forest* menunjukkan bahwa akurasi terbaik dilakukan oleh SVM dan NBC yakni sebesar 74.26% dan *Random Forest* memiliki akurasi yang sama sebesar 63.97% Adapun data komentar yang didapat didapatkan selama 1 tahun terakhir sebanyak 1945 komentar dalam rentang waktu Januari 2023 sampai dengan Desember 2023. Dari 1945 data komentar menunjukkan terdapat komentar positif sebanyak 531 komentar, komentar netral sebanyak 868 komentar, dan komentar negative sebanyak 546 komentar. Selanjutnya Hasil analisis *Social CRM* menunjukkan bahwa pelanggan akan tetap bertahan di perusahaan.

REFERENCES

- [1] M. Jami Pour and M. Hosseinzadeh, *An integrated framework of change management for social CRM implementation*, vol. 19, no. 1. Springer Berlin Heidelberg, 2021. doi: 10.1007/s10257-020-00479-z.
- [2] P. Harrigan, M. P. Miles, Y. Fang, and S. K. Roy, "The role of social media in the engagement and information processes of social CRM," *Int. J. Inf. Manage.*, vol. 54, no. April, p. 102151, 2020, doi: 10.1016/j.ijinfomgt.2020.102151.
- [3] P. Charoensukmongkol and P. Sasatanun, "Social media use for CRM and business performance satisfaction: The moderating roles of social skills and social media sales intensity," *Asia Pacific Manag. Rev.*, vol. 22, no. 1, pp. 25–34, 2017, doi: 10.1016/j.apmr.2016.10.005.
- [4] R. B. Christoph F. Breidbach and L. Hollebeek, "The service revolution and its marketing implications: service logic vs service-dominant logic," *Manag. Serv. Qual.*, vol. 24, no. 6, pp. 592–611, 2014.
- [5] V. Guerola-Navarro, H. Gil-Gomez, R. Oltra-Badenes, and J. Sendra-Garcia, "Customer relationship management and its impact on innovation: A literature review," *J. Bus. Res.*, vol. 129, no. March, pp. 83–87, 2021, doi: 10.1016/j.jbusres.2021.02.050.
- [6] J. P. Wildyaksanjani and D. Sugiana, "Strategi Customer Relationship Management (CRM) PT Angkasa Pura II (Persero)," *J. Kaji. Komun.*, vol. 6, no. 1, p. 10, 2018, doi: 10.24198/jkk.v6i1.8754.
- [7] R. Perez-Vega, P. Hopkinson, A. Singhal, and M. M. Mariani, "From CRM to social CRM: A bibliometric review and research agenda for consumer research," *J. Bus. Res.*, vol. 151, no. June, pp. 1–16, 2022, doi: 10.1016/j.jbusres.2022.06.028.
- [8] R. S. Bahri and L. Lahindah, "Cross Channel Integration Dalam Meningkatkan Retensi Pelanggan Pada Industri Ritel," *Ekonomi, Keuangan, Investasi dan Syariah*, vol. 3, no. 3, pp. 495–501, 2022, doi: 10.47065/ekuitas.v3i3.1220.
- [9] S. W. Sari, S. Sunaryo, and M. Mugiono, "the Effect of Service Quality on Customer Retention Through Commitment and Satisfaction As Mediation Variables in Java Eating Houses," *J. Apl. Manaj.*, vol. 16, no. 4, pp. 593–604, 2018, doi: 10.21776/ub.jam.2018.016.04.05.
- [10] A. P. Sutrisno and I. D. Mayangsari, "Pengaruh Penggunaan Media Sosial Instagram @Humasbdg Terhadap Pemenuhan Kebutuhan Informasi Followers," *J. Common*, vol. 5, no. 2, pp. 118–133, 2022, doi: 10.34010/common.v5i2.5143.
- [11] A. N. Jalal, M. Bahari, and A. K. Tarofder, "Transforming traditional CRM into social CRM: An empirical investigation in Iraqi healthcare industry," *Heliyon*, vol. 7, no. 5, p. e06913, 2021, doi: 10.1016/j.heliyon.2021.e06913.
- [12] T. Lohanda and A. Rusdianto Berto, "Dapatkah Aktivitas Manajemen Hubungan Pelanggan Melalui Media Sosial Meningkatkan Loyalitas Pelanggan? Can Social Customer Relationship Management Activities Evoke Customer Loyalty?," *J. Stud. Komun. dan Media*, vol. 25, no. 2, pp. 267–276, 2021, doi:



- 10.31445/jskm.2021.4439.
- [13] J. Latuny, “Analysis Social CRM against Customer Retention Using Naive Bayes Classifier (Case Study: Xyz. Ltd),” *Int. J. Comput. Sci. ...*, vol. 7, no. 7, pp. 148–157, 2018, [Online]. Available: <https://search.proquest.com/openview/5b4cec6c4478c2edafb523ad2ad6275a/1?pq-origsite=gscholar&cbl=2044552>
- [14] R. Sanusi, F. D. Astuti, and I. Y. Buryadi, “Sentiment analysis on twitter towards pre-employment card program with recurrent neural network,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 89–99, 2021.
- [15] S. Sirisuriya, “A Comparative Study on Web Scraping,” *8th Int. Res. Conf. KDU*, no. November, pp. 135–140, 2015.
- [16] M. Harahap, B. P. A. Sihombing, O. A. F. Laia, B. T. Saragih, and K. Dharma, “Analisis Sentimen Review Penjualan Produk Umkm Pada Kabupaten Nias Dengan Komparasi Algoritma Klasifikasi Machine Learning,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 5, no. 2, pp. 147–154, 2021, doi: 10.46880/jmika.vol5no2.pp147-154.
- [17] N. Hafidz, S. Anggraeni, and W. Gata, “Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes,” 2019.
- [18] M. Naïve and B. Classifier, “Mega kurnia maulidina 5150411382,” 2020.
- [19] Dedi Darwis, Nery Siskawati, and Zaenal Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *J. TEKNO KOMPAK*, vol. 15, no. 1, pp. 131–145, 2020.
- [20] S. Lamrhari, H. Elghazi, and A. El Faker, “Random Forest-based Approach for Classifying Customers in Social CRM,” *2020 IEEE Int. Conf. Technol. Manag. Oper. Decis. ICTMOD 2020*, 2020, doi: 10.1109/ICTMOD49425.2020.9380602.
- [21] D. Pramana, M. Afdal, Mustakim, and I. Permana, “Analisis Sentimen Terhadap Pindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors,” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1306–1314, 2023, doi: 10.30865/mib.v7i3.6523.
- [22] R. Sistem, S. Kurniawan, W. Gata, D. A. Puspitawati, M. Tabrani, and K. Novel, “Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada,” vol. 1, no. 10, pp. 2–8, 2021.
- [23] M. H. Saragih and A. S. Girsang, “Sentiment analysis of customer engagement on social media in transport online,” *Proc. - 2017 Int. Conf. Sustain. Inf. Eng. Technol. SIET 2017*, vol. 2018-Janua, pp. 24–29, 2017, doi: 10.1109/SIET.2017.8304103.
- [24] P. S. M. Suryani, L. Linawati, and K. O. Saputra, “Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia,” *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, p. 145, 2019, doi: 10.24843/mite.2019.v18i01.p22.
- [25] W. S. Gemilang, “Artificial Neural Networks - Models and Applications,” *Artif. Neural Networks - Model. Appl.*, vol. 7, no. 4, pp. 2849–2855, 2016, doi: 10.5772/61493.
- [26] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm,” *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [27] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, “Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes , Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja,” vol. 5, no. 4, pp. 578–584, 2021.
- [28] S. F. Sabbeh, “Machine-learning techniques for customer retention: A comparative study,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 273–281, 2018, doi: 10.14569/IJACSA.2018.090238.
- [29] F. Komunikasi, U. M. Surakarta, J. A. Yani, and T. Pos, “Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest,” vol. 9, no. 1, 2017.