

Implementasi Machine Learning Dengan Metode Text Mining Pada Twitter

Muhamad Ryansyah¹, Kudiantoro Widiyanto², Hamdun Sulaiman^{3*}, Sidik⁴, Andria Nugraha⁵

^{1,5}Program Studi Sistem Informasi, Universitas Nusa Mandiri

²Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika

³Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

⁴Program Studi Informatika, Universitas Nusa Mandiri

*hamdun.hsl@bsi.ac.id

Abstrak

Saat ini PT. Telkom Indonesia (Indihome), menggunakan peran media sosial sebagai bentuk kepedulian terhadap pelanggannya untuk menangani keluhan. *Tweet* dari pelanggan indihome pada sosial media *twitter* di tangani oleh divisi *customer service* Indihome. Manualnya proses pengkategorian yang dilakukan oleh divisi *customer service* indihome pada setiap narasi *tweet* keluhan "*complaint*" yang masuk ke *twitter @indihome*, membuat proses tersebut dinilai tidak efisien. Tujuan penelitian adalah untuk memberikan solusi terkait permasalahan pengkategorian *tweet* keluhan (*complaint*) dan membuatkan *tools* yang dapat melakukan ekstraksi terhadap narasi keluhan "*complaint*" *tweet* berbahasa Indonesia. Metode penelitian yang digunakan adalah komparatif. Disisi lain, *tools gataframework* dan *rapidminer* digunakan juga dalam penelitian ini untuk membantu *preprocessing* dan *cleansing* pada *dataset* hingga membantu menciptakan *corpus* dan *sentiment analysis*. Total *dataset* setelah dilakukan *cleansing* dan *preprocessing* menjadi 1.510. Berdasarkan metode yang diusulkan dalam penelitian ini pada algoritma klasifikasi *Support Vector Machine*, Kategori tertinggi didapati memiliki tingkat *accuracy* 82,42%, *precision* 75,33%, dan *recall* 98,75% dengan *AUC* sebesar 0,826.

Kata kunci: Complaints, Indihome , Text Mining, Twitter

Abstract

Currently PT. Telkom Indonesia (Indihome), uses the role of social media as a form of concern for its customers to handle complaints. Tweets from indihome customers on social media twitter are handled by the customer service division of Indihome. The manual of the categorization process carried out by the customer service division of Indihome on every narration of the "complain" complaint tweet that goes to @indihome twitter, makes the process considered inefficient. The purpose of this research is to provide solutions related to the problem of categorizing complaint tweets and to develop tools that can extract the narration of "complain" tweets in Indonesian. The research method used is comparative. On the other hand, gataframework and rapidminer tools are also used in this research to assist in preprocessing and cleaning of datasets to help create corpus and sentiment analysis. The total dataset after cleansing and preprocessing is 1,510. Based on the method proposed in this study on the Support Vector Machine classification algorithm, the highest category was found to have 82.42% accuracy, 75.33% precision, and 98.75% recall with an AUC of 0.826.

Keywords: Complaints, Indihome , Text Mining, Twitter

1. Pendahuluan

Indihome merupakan sebuah produk layanan komunikasi dari PT. Telekomunikasi Indonesia TELKOM. Indihome yang menyediakan beberapa

paket layanan komunikasi seperti data internet, telepon rumah dan tv kabel bagi masyarakat. Dalam menyediakan paket layanan komunikasi tentunya terdapat hambatan dan kendala yang

dialami sehingga muncul suatu penilaian sentimen dari masyarakat baik itu positif maupun negatif terhadap Indihome. Untuk itu dalam penelitian ini akan membahas mengenai seberapa besar sentimen yang diberikan oleh masyarakat kepada Indihome. Dari hasil sentimen positif maupun negatif yang muncul nantinya sebagai bahan evaluasi Indihome untuk menjawab tantangan di masa depan.

Dengan manualnya proses yang dilakukan oleh divisi customer service Indihome untuk melakukan pengkategorian setiap narasi tweet keluhan “complaint” yang masuk ke twitter @indihome, maka saat ini PT. Telkom Indonesia (Indihome) memerlukan sebuah tools yang dapat melakukan ekstraksi terhadap narasi keluhan “complaint” tweet berbahasa Indonesia. Berdasarkan hal tersebut metode text mining dapat membantu melakukan pendeteksian kategori setiap narasi tweet. Dengan pemanfaatan text mining salah satu diantaranya adalah sebagai classifier yang dapat membantu mengklasifikasi data yang dapat meningkatkan sistem operasi dalam proses pencarian karena data telah dikelompokkan berdasarkan kategori tertentu secara signifikan menggunakan proses machine learning secara supervised maupun unsupervised [1].

Analisis Sentimen juga dikenal sebagai opinion mining, adalah proses otomatis untuk memahami, mengekstraksi, dan memproses data tekstual

untuk mendapatkan informasi sentimental yang terkandung dalam sebuah kalimat, baik opini itu positif maupun negatif [2]. Masalah yang dihadapi oleh setiap bisnis adalah bagaimana memuaskan pelanggan sesuai dengan harapan perusahaan dan keinginan pelanggan setelah membuat perbandingan kinerja atau hasil yang dia rasakan terhadap harapannya [3].

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Berkaitan dengan penelitian analisis sentiment text mining, tentunya sudah banyak yang meneliti perihal tersebut, terdapat 5 penelitian terkait:

- Penelitian yang dilakukan oleh Felix Fridom Mailo, Luffan Lazuardi “Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia” yang terbit pada tahun 2019. Penelitian tersebut menghasilkan analisis sentimen tweet didapatkan sentimen positif sebanyak 22.246 (51,2%) tweet, diikuti oleh sentimen negatif sebanyak 12.015 (27,7%) tweet dan sentimen netral dengan jumlah 9.174 (21,1%) dari total 43.435 tweet. Nilai akurasi dengan algoritma Naïve Bayes didapatkan hasil nilai akurasi sebesar 94% [15].
- Penelitian selanjutnya dilakukan oleh Faisal Al isfahani , dan Rifki Mubarak dengan judul “Analisis Sentimen Pengguna Twitter

Terhadap Kebijakan Pemberlakuan Pembatasan Sosial Berskala Besar (Psbb) Dengan Metode Naïve Bayes” yang terbit pada tahun 2021. Penelitian tersebut menghasilkan Tweet terkait opini PSBB diunduh secara otomatis melalui tools twitter scraper, preprocessing dan klasifikasi menggunakan metode naive bayes. Hasil percobaan pada penelitian dari data twitter terkait PSBB diperoleh sentiment positif, negatif dan neutral dengan nilai akurasi rata 58.2% [16].

- Penelitian selanjutnya yaitu dari Syiva Multi Fani , Rukun Santoso , Suparti dengan judul “Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering” yang terbit pada tahun 2021. Pada penelitian ini dilakukan clustering menggunakan software R. Metode clustering yang digunakan adalah metode K-means clustering untuk mengelompokkan data tekstual berdasarkan kesamaan konten yang dimiliki ke dalam beberapa cluster. Jumlah anggota masing-masing cluster kurang merata. Cluster 1 memiliki anggota sebanyak 827, cluster 2 sebanyak 15, cluster 3 sebanyak 41, cluster 4 sebanyak 5, cluster 5 sebanyak 46, dan cluster 6 sebanyak 28. Cluster 1 memiliki paling banyak anggota daripada cluster yang lain karena perbedaan

bobot yang cukup jauh dan banyaknya kata yang sama terdapat pada cluster 1 seperti kata “diskon” dan “promo” yang jumlahnya sangat banyak. Sebagian besar tweet yang ditulis oleh akun @blibliidotcom juga memuat tentang aktivitas perbelanjaan yang termasuk dalam cluster 1 [17].

- Penelitian selanjutnya dari Tri Putri Lestari dengan judul “Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA)” yang terbit pada tahun 2022. Berdasarkan hasil penelitian dan pengujian metode Support Vector Machine untuk mengklasifikasi pinjaman online dengan tingkat nilai Accuracy sebesar 86.6%, dengan precision sebesar positif sebesar 86 %, netral 1.00% dan negatif sebesar 87%, recall positif sebesar 90%, netral 87% dan negatif sebesar 26% dan F1-Score positif sebesar 88% netral 42% dan negatif sebesar 86%. Kemudian pada tahap Social Network Analysis terdapat akun paling berpengaruh yaitu influencer @alvinline21 dengan 1402 nodes [18].
- Penelitian yang kelima dilakukan oleh Aditiya Hermawan, Indrico Jowensen, Junaedi, Edy dengan judul “Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine” terbit pada tahun 2023. Penggunaan text mining

menggunakan SVM dalam melakukan klasifikasi pada tweet berbahasa Indonesia mempunyai akurasi 73% berdasarkan pada 10 kali percobaan yang dilakukan dengan keyword dan waktu yang berbeda-beda. Kemudian nilai presisi yang didapatkan adalah 67% dan nilai recall yang didapatkan 54% [19].

2.2. Landasan Teori

1. Keluhan Pelanggan

Keluhan pelanggan merupakan ketidakpuasan yang didasarkan oleh pelanggan ketika membeli dan menggunakan barang, baik itu berupa bentuk produk fisik atau jasa pelayanan. Keluhan pelanggan tidak dapat diabaikan karena membuat pelanggan merasa tidak dihargai dan tidak didukung sama sekali [6]. Oleh karena itu dapat disimpulkan bahwa keluhan pelanggan adalah tanggapan dari pelanggan yang dikirim ke perusahaan yang cenderung negatif, dapat tertulis atau lisan [6].

2. Twitter

Twitter adalah layanan microblogging online yang memungkinkan pengguna untuk berbagi konten. Twitter adalah layanan perpesanan yang membagikan berbagai karakteristik dengan alat komunikasi yang sudah digunakan. Twitter memiliki unsur-unsur yang mirip dengan surat elektronik (email), instant messenger, pesan

singkat (SMS), blogging, RSS, jejaring sosial, dan sebagainya [7].

3. Text mining

Text mining adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data[8].

4. Data Mining

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambil keputusan diwaktu yang akan datang, pola-pola ini dikenali perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lain [9].

5. Analisis Sentimen

Analisis sentimen adalah kombinasi dari text mining dan data mining, atau metode yang digunakan untuk menangani pendapat yang berbeda dari konsumen atau pakar melalui berbagai cara yang berbeda, yang berkaitan dengan produk, layanan, atau agen . Analisis sentimen adalah metode yang digunakan untuk memahami, mengekstrak data opini, dan secara otomatis memproses data tekstual untuk mendapatkan emosi yang terkandung dalam opini. Analisis Sentimen mencakup 3 jenis opini, yaitu opini positif, opini negatif dan opini netral. Jadi, dengan perusahaan analisis sentimen atau lembaga terkait, tanggapan masyarakat terhadap layanan atau produk dapat ditemukan, baik melalui umpan balik komunitas atau pakar [10].

6. Naïve Bayes

Naive bayes memiliki kelebihan, menghitung cepat, algoritma sederhana dan memiliki akurasi tinggi [11]. Metode Naïve Bayes adalah metode yang menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes [12].

7. Machine Learning

Kecerdasan Buatan merupakan salah satu bidang dalam ilmu komputer yang ditujukan pada pembuatan software dan hardware yang dapat berfungsi sebagai sesuatu yang dapat berpikir seperti manusia. Kecerdasan buatan banyak

digunakan untuk memecahkan berbagai masalah seperti bisnis robotika, bahasa alami, matematika, game, persepsi, diagnosis medis, teknik, analisis keuangan, analisis sains, dan penalaran [13].

8. Gata Framework

Gata Framework adalah software berbasis web yang menyediakan text pre- processing yang mengakses dalam membantu proses stemming. Gata Framework bahasa indonesia dapat diakses pada halaman gataframework.com/textmining [14].

9. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE adalah metode oversampling dimana data pada kelas minoritas direplikasi menggunakan data agregat yang diperoleh dengan menduplikasi data pada kelas minoritas. Oversampling di SMOTE mengambil instance dari kelas minoritas, kemudian menemukan ketetangga terdekat dari setiap kasus dan kemudian menghasilkan instance komposit alih-alih menduplikasi instance kelas minoritas. Oleh karena itu, dapat menghindari masalah overfitting [14]

3. Metode Penelitian

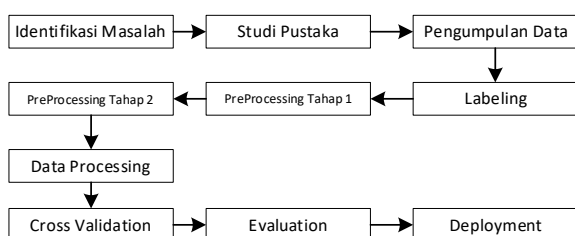
3.1. Metode Pengumpulan Data

Pada penelitian ini diambil dari twitter menggunakan 250 data per hari selama 7 hari yang diambil dengan kata yang berkaitan keluhan

pelanggan yang berkaitan dengan layanan Indihome dengan menggunakan *software rapidminer*.

3.2. Tahap Penelitian

Dibawah ini adalah tahapan pada text mining menggunakan metode naive bayes classifier pada twitter @indihome seperti pada gambar berikut:



Gambar 1. Tahap Penelitian

3.3. Metode Penelitian

Dalam implementasi machine learning dalam penelitian ini, dilakukan:

1. Observasi

Dilakukan observasi dengan mengamati interaksi akun *twitter* @indihome pada pelanggan Indihome di media sosial *twitter*.

2. Studi Pustaka

Studi pustaka dilakukan untuk mengumpulkan lebih banyak sumber informasi yang relevan dan untuk perbandingan dalam penelitian ini.

3. Pengumpulan Data

Pengumpulan data yang dilakukan pada penelitian ini yaitu dengan melakukan crawling (penarikan data) dari twitter menggunakan *rapidminer*

3.4. Analisis Data

Media sosial menjadi salah satu sarana yang digunakan pengguna internet untuk mengakses, berbagi, dan mendiskusikan masalah obesitas. Twitter merupakan salah satu media sosial yang secara real time sering dipilih untuk mengkomunikasikan hal ini. Melalui analisis sentimen dengan metode text mining pada twitter, kita dapat memahami bagaimana orang menggambarkan dan mengekspresikan persepsi mereka tentang obesitas dengan cara yang positif, negatif, atau netral. Analisis ini penting untuk melihat sejauh mana media sosial seperti twitter saat ini digunakan sebagai sarana penyebaran informasi bidang jasa pelayanan di Indonesia. Penelitian ini bertujuan untuk mengidentifikasi analisis sentimen twitter terkait obesitas di Indonesia menggunakan metode text mining. Pada proses ini akan dilakukan dengan menggunakan algoritma naive bayes. Gata framework akan melakukan klasifikasi sentimen secara otomatis apabila dibuat dengan algoritma naive bayes untuk mengetahui apakah kalimat dari pelanggan yang masuk pada akun @indihome bernarasi sentimen "negative" apabila bersifat complaint atau "positive" apabila bukan merupakan complaint.

4. Hasil dan Pembahasan

Penelitian ini dilakukan pemahaman untuk mencari metode dengan pendekatan model

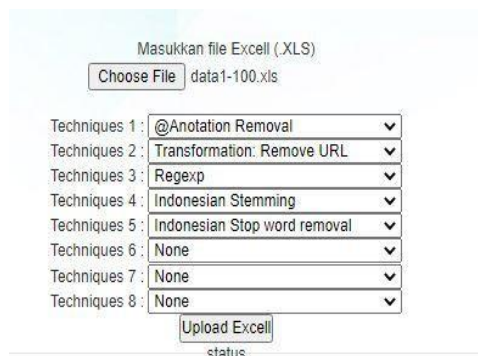
pengkategorian terbaik agar dapat membantu pada saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma.

4.1. Data Understanding

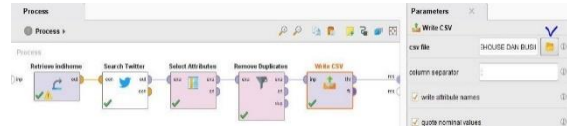
Tahap ini merupakan proses memahami data yang akan dijadikan sebagai bahan penelitian untuk dapat melangkah ke tahap selanjutnya yaitu preprocessing. Di bawah ini adalah langkah-langkah yang harus dilakukan. Siapkan data tweet yang diambil dari media sosial twitter, dimana data sebanyak 2.017 data. Proses selanjutnya adalah pembersihan data, menghapus data duplikat. Jumlah data awal sebanyak 2.017 setelah dibersihkan dengan menghapus duplikat menjadi 1.500 data.

4.2. Data Preparation

Langkah kedua adalah menyiapkan data untuk apa yang disebut langkah text preprocessing, dengan menggunakan dua aplikasi preprocessing. Pertama, gunakan Gataframework yang dapat diakses melalui tautan www.gataframework.com dan menggunakan preprocessing software rapidminer.



Gambar 2. Model *Preprocessing Data Local* menggunakan *Gataframework*



Gambar 3. Model *Preprocessing Data Local* menggunakan *Rapid Miner*

Gambar 3 untuk memastikan tidak ada data duplikat, maka agar data dapat disimpan ke komputer lokal di menu operator, kita pilih "write CSV". Pada pengaturan file csv, isikan di mana file harus disimpan dalam folder.

4.3. Labeling Data

Tabel 1. Desain Tabel Model Labeling Data

1	@IndiHomeCare @TelkomCare Belum Bisa	Negatif
2	@IndiHomeCare @TelkomCare Sdh di DM lg dr 7jam yang lalu	Negatif
3	@InduVaya @TelkomCare Hello	Positif
4	@HAKU_sugi @TelkomCare Hello	Positif

Tabel 1 adalah tabel labeling data dengan sentimen *positive* atau *negative*

4.4. Annotation Removal

Tabel 2. Desain Tabel Model Annotation Removal

Text;Id	Sentimen (Positif / Negatif)	Transformation: Remove URL
@IndiHomeCare @TelkomCare Sdh di DM lg dr 7jam yang lalu	Negatif	sdh di dm lg dr 7jam yang lalu

Proses analisis mengurai teks berdasarkan "white space", di mana semua anotasi terkandung dalam "tweet" akan dihapus dan semua huruf besar diubah menjadi huruf kecil.

4.5. Transformation Remove URL

Tabel 3. Desain Tabel Model *Transformation Remove URL*

Text;Id	Sentimen (Positif / Negatif)	Transformation: Remove URL
@TelkomCare wifi koq lelet ya?	Negatif	wifi koq lelet ya?

Setelah dari proses anotasi removal, kemudian dilanjutkan ke proses transformation remove URL, pada proses ini link atau URL yang terkandung pada tweet akan dihilangkan.

4.6. Tokenize Regular Expression

Tabel 4. Desain Tabel Model Tokenize Regular Expression (Regexp)

TextId	Sentimen (Positif / Negatif)	Regexp
@BMarubeni @TelkomIndonesia @TelkomCare @Berickthohir Update tanggal 22 oktober 2021 complain disini aja udah 9 hari ga ada perubahan	Negatif	update tanggal oktober complain disini aja udah hari ga ada perubahan

Hasil Transformation Remove URL dilanjutkan melalui proses tokenization (Regexp) dimana semua kata dalam setiap dokumen dikumpulkan dan tanda baca dihilangkan, dan jika ada simbol, karakter khusus atau apapun yang bukan huruf.

4.7. Indonesian Stopword Removal

Tabel 5. Desain Tabel Model Indonesian Stopword Removal

TextId	Sentimen (Positif / Negatif)	Indonesian Stop word removal
Bang @IndiHome @TelkomCare kenapa gangguan masal tidak selesai sudah 2 hari? Ngga ada kompensasi kalo begini? Apakah ada tiang tumbang karena #SalamDariBinjai ? Serius aku bang	Negatif	bang ganggu masal selesai kompensasi kalo tiang tumbang salamdariBinjai serius bang






Hasil tokenization akan dilanjutkan pada proses Indonesian stopword removal. Pada proses ini kata-kata yang tidak relevan akan dihilangkan, seperti kata tetapi, dengan, kata tidak memiliki arti tersendiri. kata lain dan tidak terkait dengan kata sentiment.

4.8. Evaluasi

AUC memiliki potensi untuk secara signifikan meningkatkan konvergensi empiris eksperimen karena memisahkan kinerja prediktif dari kondisi operasi dan merupakan ukuran prediktif umum.

Selanjutnya, AUC memiliki interpretasi yang jelas secara statistik

Tabel 6. Symbol Nilai AUC

AUC	Meaning	Symbol
0.90 – 1.00	<i>excellent classification</i>	
0.80 – 0.90	<i>good classification</i>	
0.70 – 0.80	<i>fair classification</i>	
0.60 – 0.70	<i>poor classification</i>	
< 0.60	<i>failure</i>	

Tabel 7. Hasil tahap preprocessing

Kriteria	Nilai
Accuracy	82.42
Precision	75.33
Recall	98.75
AUC	0.826

4.9. Deployment

Dari hasil pengujian melalui tahapan eksperimen dan evaluasi dari model yang akan digunakan, akan menghasilkan nilai-nilai yang dapat digunakan sebagai data acuan dalam pembuatan model implementasi *machine learning*. Deployment dilakukan melalui *link* www.gataframework.com maka akan tampil halaman *login* seperti pada gambar dibawah ini:



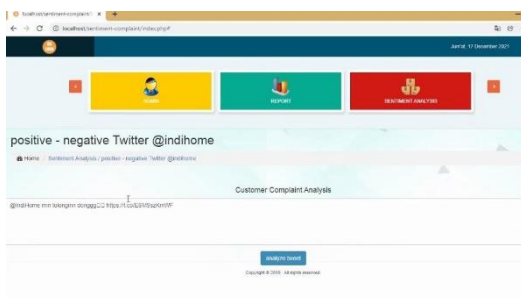
Gambar 4. Tampilan Login

Pada gambar dibawah ini geser menu ke sebelah kanan dengan mengklik tanda panah “ > “ lalu klik “*sentiment Analysis*” kemudian klik “*positive-negative twitter @indihome*”



Gambar 5. Tampilan Sentiment Analyst

Klik “Analyze Tweet” untuk menganalisis kalimat customer complaint yang diambil secara otomatis dari twitter langsung seperti gambar dibawah ini:



Gambar 6. Tampilan Postive - Negative Twitter @indihome



Gambar 7. Tampilan bobot kategori

Pada gambar 7, setelah data tweet dilakukan analyze tweet maka, Hasil dari perhitungan untuk kategori tersebut menghasilkan kesimpulan negative.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai permasalahan pengkategorian tweet dengan narasi “complaint” atau keluhan pelanggan PT. Telkom Indonesia, Tbk di sosial media twitter dengan menggunakan mention @indihome dapat ditarik kesimpulan yaitu pendekatan dengan menggunakan metode text mining. Berdasarkan metode yang diusulkan dalam penelitian ini pada algoritma klasifikasi Support Vector Machine, kategori tertinggi didapati memiliki tingkat accuracy 82,42%, precision 75,33%, dan recall 98,75% dengan AUC sebesar 0,826. Dengan menggunakan machine learning, perusahaan dapat mengetahui informasi atas masalah dan hambatan operasional sehingga perusahaan

dapat meninjau dan segera mendapatkan solusi untuk ditindak lanjuti demi terciptanya kepuasan pelanggan.

6. Daftar Pustaka

- [1] K. Sosial and D. A. N. Kinerja, "Metadata, citation and similar papers at core.ac.uk 1," vol. 2, no. 1, pp. 1–118, 2009.
- [2] Novianti, "Pengaruh Kualitas Jaringan Internet Terhadap Kepuasan Pelanggan Indihome," pp. 1–59, 2020.
- [3] A. V. Sudiantoro *et al.*, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier," *Din. Inform.*, vol. 10, no. 2, pp. 398–401, 2018.
- [4] I. A. Paramitha, "Tinjauan Pustaka Tinjauan Pustaka," *Conv. Cent. Di Kota Tegal*, pp. 6–37, 2017.
- [5] M. Kurniasih, "Bab li Landasan Teori," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 8–24, 2018.
- [6] S. R. I. Rezeki, "Penggunaan Sosial Media Twitter dalam Komunikasi Organisasi (Studi Kasus Pemerintah Provinsi Dki Jakarta Dalam Penanganan Covid-19)," *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020.
- [7] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [8] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid – 19 Menggunakan Metode Naïve," *Jurnal Sains dan Informatika*, vol. 7, no. 2, p. 155 – 164 , 2021.
- [9] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *Jurnal Teknik Komputer*, vol. 4, no. 1, p. 156 – 161 , 2018.
- [10] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [11] R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31311/ji.v6i1.4685.
- [12] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [13] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.
- [14] Dewani, Andini and Aulia, "Pengaruh Kualitas Jaringan IndiHome Terhadap Customer Experience," *Semin FORTEI 2019*, no. 1, pp. 67-72, 2019
- [15] Mailo, F. F., & Lazuardi, L. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. *Journal of Information Systems for Public Health*, 28-36.
- [16] isfahani, F. A., & Mubarak, R. (2021). Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Sosial Berskala Besar (Psbb) Dengan Metode Naïve Bayes. *Jurnal Siliwangi Seri Sains dan Teknologi*, 19-24.

- [17] Fani, S. M., Santoso, R., & Suparti. (2021). Penerapan text mining untuk melakukan clustering data tweet akun bibli pada media sosial twitter menggunakan k-means clustering. *Jurnal gaussian*, 583-593.
- [18] Lestari, T. P. (2022). Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA). *Jurnal Informatika Ekonomi Bisnis*, 65-71 .
- [19] Hermawan, A., Jowensen, I., Junaedi, & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *Jurnal Sains dan Teknologi*, 129-137.