Terakreditasi SINTA Peringkat 4
Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 158/E/KPT/2021
masa berlaku mulai Volume 5 Nomor 2 Tahun 2018 sampai Volume 10 Nomor 1 Tahun 2023

Terbit online pada laman web jurnal: **https://jurnal.plb.ac.id/index.php/tematik/index**

TEMATIK

Jurnal **Teknologi Informasi Komunikasi (e-Journal)**

Vol. 10 No. 2 (2022) 246 - 251 ISSN Media Elektronik: 2443-3640

Analisis Perbandingan Sentimen Pengguna Twitter Terhadap Layanan Salah Satu Provider Internet Di Indonesia Menggunakan Metode Klasifikasi

Comparative Analysis of Twitter User Sentiments Towards the Services of One Internet Provider in Indonesia Using the Classification Method

Della Puspita Sari[1], Budiman[2], Nur Alamsyah[3*]
[1]Sistem Informasi, Fakultas Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia
[1]dellaps040@gmail.com, [2]budiman@unibi.ac.id, [3]nuralamsyah@unibi.ac.id

Abstract

The Internet is needed for everyday life, whereas in Indonesia there are many internet service providers, one of which is indihome. Sentiment analysis itself aims to classify a text into Negative, Positive and Neutral classes. On the twitter platform, there are many reviews about internet providers, one of which is indihome, because of poor service or just to appreciate the services provided. Based on the calculation of the results obtained 71.1% negative, 21.1% positive and 7.7% neutral. The data obtained is not balanced, therefore the classification process is assisted using Smote. The results of the comparison of the four methods used are Support Vector Machine, Naïve Bayes, Random forest, Decision tree. From the overall comparison, the highest accuracy without smote or using smote is Support Vector Machine with an accuracy level of 89% AUC level of 89% if using smote gets 93% accuracy and 97% AUC level with 80% training data and 20% testing.

Keywords: smote, support vector machine, naïve bayes, random forest, decision tree

Abstrak

Internet sangat dibutuhkan untuk kehidupan sehari-hari, dimana di indonesia sudah banyak penyedia layanan internet salah satunya indihome. Analisis sentimen sendiri bertujuan untuk mengelompokan suatu teks kedalam kelas Negative, Positive dan Netral. Pada platform twitter banyak sekali ulasan mengenai provider internet salah satunya indihome, karena adanya pelayanan yang kurang baik atau untuk sekedar mengapresiasi pelayanan yang diberikan. Berdasarkan perhitungan hasil yang didapatkan 71,1% negative, 21,1% positive dan 7,7% netral. Data yang diperoleh tidak seimbang maka dari itu proses pengklasifikasian di bantu menggunakan Smote. Hasil perbandingan keempat metode yang digunakan ialah Support Vector Machine, Naïve Bayes, Random forest, Decision tree. Dari keseluruhan perbandingan accuracy paling tinggi tanpa smote ataupun menggunakan smote ialah Support Vector Machine dengan tingkat accuracy 89% tingkat AUC 89% jika menggunakan smote mendapatkan accuracy 93% dan tingkat AUC 97% dengan data training 80% dan testing 20%.

Kata kunci : smote, support vector machine, naïve bayes, random forest, decision

1. Pendahuluan

Internet adalah sebuah jaringan yang dimana bisa saling terhubung satu sama lain. Internet juga sering digunakan untuk bermain media sosial, media sosial merupakan tempat untuk menyampaikan berbagai macam kritik dan saran [1]. Salah satu contoh media sosial yang digunakan ialah twitter, twitter merupakan salah satu media sosial yang memungkinkan penggunanya untuk menulis tentang berbagai topik dan membahas isu-isu yang sedang terjadi [2]. Jaringan internet saat ini sangatlah luas, mengingat di Indonesia sendiri banyak sekali penyedia provider internet untuk kebutuhan sehari-hari [3]. Dianalisis dari flip.id ada tujuh provider tercepat penyedia layanan internet diantaranya ialah Indihome, First Media, CBN, Transvision, My Republic, Biznet. Sebenarnya banyak sekali saat ini provider baru, seperti MNC, Iconnect dari PLN.

Menurut survey APJII, per tahun 2022 Indihome menjadi provider yang banyak digunakan oleh

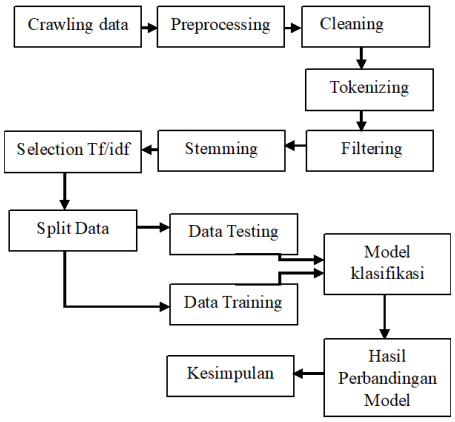
Diterima Redaksi: 27-10-2023 | Selesai Revisi: 16-12-2023 | Diterbitkan Online: 16-12-2023

masyarakat Indonesia. IndiHome merupakan perusahaan telekomunikasi indonesia milik PT. telekomunikasi indonesia mencakup 3 layanan biasa dikenal dengan Internet wifi, Indihome tv dan telpon rumah [4]. Analisis sentimen adalah kegiatan yang digunakan untuk menganalisis pendapat atau opini seseorang tentang suatu topik. Tugas dasar Analisis sentimen adalah mengklasifikasikan beberapa teks dari dokumen, kalimat atau fitur, kalimat dari fitur tersebut bisa bersifat positif, negatif dan netral [5]. Di indonesia banyak sekali yang menggunakan provider indihome maka pasti akan banyak sekali ulasan mengenai pelayanan yang diberikan oleh pihak indihome kepada customer [6]. Ulasan yang diberikan oleh customer indihome berbagai macam mulai dari hal yang menyenangkan ataupun tidak menyenangkan. Ulasan tersebut akan menjadi landasan untuk penelitian ini [7]. Melakukaan analisis sentimen, diperlukan suatu metode klasifikasi & seleksi fitur yang mempuni agar didapatkan hasil akurasi yang maksimal, pada penelitian [8], sentiment analisis yang dilakukan hanya menggunakan model utama yaitu naïve bayes dan tidak membandingkan dengan model lainnya . Dalam penelitian ini menggunakan metode Naïve Bayes, Decision Tree, Random forest dan Support Vector Machine untuk mengklasifikasikan negative, posistive dan netral untuk memberikan optimasi nilai yang baik menggunakan SMOTE (Synthetic Minority Oversampling Technique) serta membandingkan metode mana yang cukup baik dengan melihat hasil AUC (Area Under Curve) [9], menggunakan python dan crawling data menggunakan rapidminer dengan query layanan indihome pada rapidminer [10].

2. Metode Penelitian

Metode penelitian ini diawali dengan mencari beberapa sumber yang relevan dengan peneletian yang dijadikan sebagai pedoman penulisan, salah satu karya ilmiah yang digunakan sebagai pedoman dengan judul artikel Analisis Sentimen terhadap Layanan Indihome di Twitter dengan Metode Machine Learning. Disimpulkan bahwa metode yang terbaik adalah metode Random Forest sebab Random Forest menghasilkan akurasi yang lebih tinggi dibandingkan dengan metode Gradient Boosting seperti terlihat pada Gambar 1 [11].

Hasil yang diperoleh sudah dilakukan cleaning dan tahapan selanjutnya menyimpan hasil dari rapidminer ke file csv, data dari rapidminer masih belum sempurna dan belum diberikan label, masih banyak noise yang ada pada kalimat tweet yang tidak ada maknanya, Langkah selanjutnya ialah menghapus beberapa noise secara manual dan memberi label pada kelas positive, negative dan netral, hasil yang sudah dilakukan cleaning dan labelling mendapatkan hasil 405 data, dan teakhir adalah hasil setelah dilakukan labelling.



Gambar 1. Metodologi Yang Digunakan

2.1. *Crawling* Data

Tahapan yang dilakukan ialah melakukan crawling data dengan menggunakan aplikasi rapidminer [12]. Aplikasi rapidminer ini harus menggunakan api twitter, alangkah lebih baiknya saat akan menggunakan rapidminer bisa membuat akun lain terlebih dahulu agar tidak menggunakan akun utama, khawatir terjadi seseuatu yang tidak diinginkan. Seteleh melakukan koneksi antara rapidminer dan twitter selanjutkan memilih selection untuk memilih data apa yang akan ditampilkan oleh rapidminer, karena pada penulisan ini yang diperlukan hanya teks saja maka, attribute yang dipilih ialah attribute teks. Pada tahapan crawling data ini dalam rapidminer bisa melakukan tahapan cleaning juga dengan memilih attribute replace dan menghapus beberapa karakter yang mempengaruhi pada tahapan analisis sentiment. Hasil data yang dilakukan crawling sebanyak 3000 data, data yang sudah di cleaning melalui rapidminer hasil nya ada pada Tabel 1.

Tabel 1. Hasil *Crawling* Data

'Text'
2 series tadi bisa kamu streaming di Disney+ Hotstar
Kalo langganannya lewat IndiHome TV bakal lebih banyak lagi benefitnya Mantap
Sekedar mau nanya
Apa bener orang yang bisa liat hantu
Di sebut indihome
IndiHome
Saya sudah bayar dan sudah ada notif dari apps my indihome
kalau pembayaran sudah berhasil tapi kenapa di tampilan indihome nya masih blm di bayar ya
kok diem aja indihome

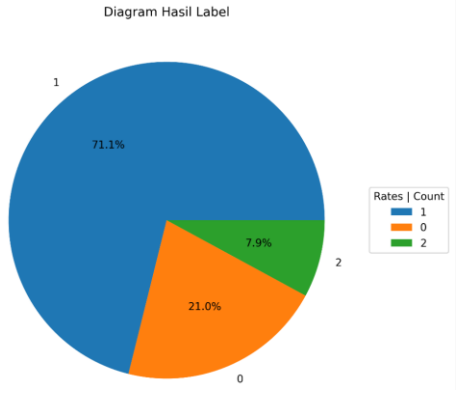
Hasil yang diperoleh sudah dilakukan *cleaning* dan tahapan selanjutnya menyimpan hasil dari rapidminer ke file csv, data dari rapidminer masih belum sempurna dan belum diberikan label, masih banyak noise yang ada pada kalimat tweet yang tidak ada maknanya, Langkah selanjutnya ialah menghapus beberapa noise secara manual dan memberi label pada kelas positive, negative dan netral, hasil yang sudah dilakukan cleaning [13]. Labelling mendapatkan hasil 405 data. Tabel 2 adalah hasil setelah dilakukan labelling.

DOI: https://doi.org/10.38204/tematik.v10i2.1578

Tabel 2. Hasil Labeling Manual

Text	Label
Kalo langganannya lewat IndiHome TV bakal lebih banyak lagi benefitnya Mantap	Positif
Saya sudah bayar dan sudah ada notif dari apps my indihome kalau pembayaran sudah berhasil tapi kenapa di tampilan indihome nya masih blm di bayar ya	Negative
betul ada indihome semua jadi mudah	Positif
fortune cookie mas dijamin mood auto lancar kek indihome	Positif
alias gue stres bgt ajg besok2 gue mau ganti ke Indihome aja	Positif
indihome lo kalo sering error begini mending kita putus aja IndiHomeCare	Negative
iya aku mau ganti ke indihome	Positif
Speedtest nya arapenta lebih cepat dari indihome	Negative

Penelitian ini dilakukan menggunakan *python*, setelah data diberikan label data akan di simpan menjadi file csv kembali, setelah itu dipanggil menggunakan alat bantu *Google Colab* untuk mengoprasikan *python*. Beberapa *library* yang digunakan di pemrograman python ialah, *"Numpy", "Pandas", "NLTK", "SlangWord", "Stopwords", "Sastrawi", "Swifter"* dan algoritma klasifikasi yang digunakan.



Gambar 2. Hasil Labelling diagram

Diagram label pada Gambar 2 menunjukan bahwa hasil dari data tersebut masih imbalance, dimana terlihat label yang sangat dominan ialah label Negative 71.2% Positif 21,1% dan netral 7,7%. Hasil yang tidak balance maka diperlukan metode tambahan agar nilai accuracy menjadi lebih baik untuk memberikan nilai optimasi yang baik maka dilakukan penambahan library smote agar menyeimbangkan data.

2.2. *Pre-Processing.*

Tahap pre-processing ini adalah untuk melakukan pembersihan dan menghapus data dari kata-kata yang tidak perlu atau tidak memiliki keterkaitan dengan sentimen, dilakukan menggunakan python tidak manual seperti diawal, ada beberpa tahapan pada tahapan ini ialah *Case folding, Cleaning, Tokenizing, Filtering, dan Stemming* [14].

Tahapan Case Folding untuk mengkonversi teks menjadi suatu bentuk standar, dimana menjadi *Lowercase* (format penulisan huruf kecil).

Tahapan Cleaning ini sudah dilakukan diawal pengambilan data, karena menggunakan alat bantu *RapidMiner* dan dilakukan secara manual untuk *filtering* dan *labelling.*

Tahapan *tokenizing* itu untuk melalukan pemotongan *string*, jadi tahapan ini untuk memecah teks sekepumpulan karakter yang terdapat dalam teks kedalam satuan kata.

Hasil proses *Tokenizing* akan dilakukan *filtering* kembali untuk pengambilan kata-kata penting hasil proses *Tokenizing*, tahapan ini membuang kata yang kurang penting atau menyimpan kata penting, ada beberapa kata yang dihapus pada tahapan ini seperti saya, dan, atau, di, ke, ok, dalam *filtering* ini hal yang digunakan ialah *Slangwords* dan *Stopwords*. Proses *slangwords* itu digunakan untuk pengecekan kata kata gaul pada dataset tersebut tahapan selanjutnya ialah proses *Stopword* untuk melakukan filter dimana membuang kata kata yang sering muncul namun tidak memberikan nilai informasi yang signifikan, dibawah ialah contoh stopwords.

Proses *stemming* dilakukan perubahan kata yang berimbuhan menjadi kata dasar dengan menghapus imbuhan di depan atau di belakang kata dasar. Tahap ini dibantu dengan *library Sastrawi* dan *Swifter* [15].

2.3. *Term Frequency/Inverse Document Frequency (TF/IDF)*

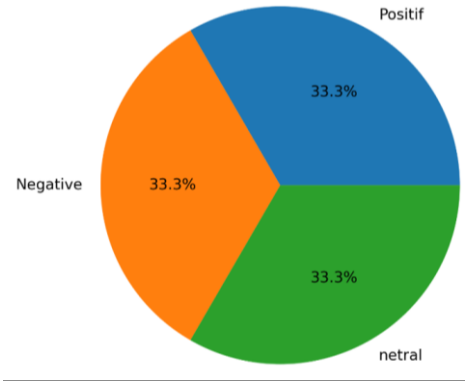
TF/IDF menghitung kemunculan dari term yang terdapat dalam dokumen. Pembuatan fitur selection untuk memudahkan dalam proses klasifikasi [16]. Pada tahapan *TF/IDF* juga penulis merubah pelebelan dari yang awalnya menggunakan string lalu dilakukan perubahan menjadi integer, dimana Negative adalah '1' Positif '0' dan Netral '2'. Tahapan ini dilakukan agar perhitungan kedepannya menjadi lebih mudah.

2.4. *Split* Data

Tahapan split data ini akan dibagi menjadi 2 data ialah data training dan data testing [17]. Pada tahapan awal penulis akan mencoba data testing 20% data training 80% kemudian akan membuat perbandingan hasil antara 70% *training* 30% *testing* dan perbandingan 50% *testing* dan 50% *training*. Tujuannya agar bisa mengetahui apakah semakin besar data testing hasil akan baik atau sebaliknya. Seteleh melakukan split data, penulis akan membuat library baru untuk data yang tidak balance menggunakan *Smote*, sesuai dengan data yang diperolah data tidak *balance*. *Smote* ini akan membuat data sintetis dari data yang minoritas dengan cara sampling ulang datanya, nantipun akan di lihat perbandingannya apakah akan merubah tingkat

DOI: https://doi.org/10.38204/tematik.v10i2.1578

accuracy nya atau tidak. Gambar 3 ini hasil dari *smote* yang dibuat [18].



Gambar 3. Hasil Smote

Hasil setelah di-smote data akan terlihat balance, kenapa 33.3% hasil keseluruhannya label yang digunakan ada 3 yaitu negative, positif dan netral. Setelah mendapatkan hasil data yang balance penulis akan melakukan klasifikasi dan akan melakukan perbandingan apakah ada pengaruh terhadap accuracy yang di dapatkan atau tidak. Setelah melakukan tahapan pre-processing dan split data ialah melakukan pengujian terhadap metode yang akan digunakan.

3. Hasil dan Pembahasan

Penelitian ini menggunakan Naïve Bayes, Support Vector Machine, Random Forest, Decision Tree, hasil dari ke empat Metode :

3.1. *Naïve Bayes*

Tabel 3 menunjukan bahwa tingkat accuracy yang tinggi ada pada data training 80% dan data Testing 20% dengan nilai accuracy tanpa smote itu di 86% dengan tingkat AUC 95% dan jika menggunakan smote di 89% accuracy serta tingkat AUC 95%.

Tabel 3. Hasil *Naïve Bayes*

Metode	Data Training	Data Testing	Precision	Recall	F1 Score	Accuracy	AUC
Naïve Bayes	80%	20%	86%	86%	86%	86%	95%
	70%	30%	85%	85%	84%	85%	89%
	50%	50%	82%	83%	82%	83%	90%
Naïve Bayes dengan Smote	80%	20%	90%	89%	89%	89%	95%
	70%	30%	87%	86%	87%	86%	92%
	50%	50%	87%	85%	86%	85%	93%

3.2. *Support Vector Machine*

Tabel 4 menunjukan bahwa tingkat accuracy yang tinggi ada pada data training 80% dan data testing 20% dengan nilai accuracy tanpa smote itu di 89% dengan tingkat AUC 89% dan jika menggunakan smote di 93% serta tingkat AUC 97%.

Tabel 4. Hasil Support Vector Machine

Metode	Data Training	Data Testing	Precision	Recall	F1 Score	Accuracy	AUC
SVM	80%	20%	90%	89%	87%	89%	89%
	70%	30%	83%	80%	76%	80%	89%
	50%	50%	85%	83%	79%	83%	89%
SVM dengan Smote	80%	20%	93%	93%	92%	93%	97%
	70%	30%	85%	85%	84%	85%	86%
	50%	50%	86%	86%	85%	86%	91%

3.3. *Random Forest.*

Tabel 5 menunjukan bahawa tingkat accuracy yang tinggi ada pada data training 70% dan data testing 30% dengan nilai accuracy tanpa smote itu di 82% dengan tingkat AUC 91% dan jika menggunakan smote di 85% serta tingkat AUC 91%.

Tabel 5. Hasil Random Forest

Metode	Data Training	Data Testing	Precision	Recall	F1 Score	Accuracy	AUC
Random Forest	80%	20%	82%	74%	69%	74%	80%
	70%	30%	77%	82%	77%	82%	91%
	50%	50%	75%	78%	73%	78%	83%
Random Forest dengan smote	80%	20%	82%	81%	79%	81%	84%
	70%	30%	86%	85%	83%	85%	91%
	50%	50%	80%	80%	77%	80%	87%

DOI: https://doi.org/10.38204/tematik.v10i2.1578

3.4. *Decision Tree.*

Tabel 6 menunjukan bahawa tingkat accuracy yang tinggi ada pada data training 50% dan data testing 50% dengan nilai accuracy tanpa smote itu di 78% dengan tingkat AUC 70% dan jika menggunakan smote di 81% serta tingkat AUC 71%.

Tabel 6 Hasil Decision Tree

Metode	Data Training	Data Testing	Precision	Recall	F1 Score	Accuracy	AUC
Decision Tree	80%	20%	74%	73%	71%	73%	67%
	70%	30%	76%	74%	75%	74%	70%
	50%	50%	78%	78%	77%	78%	70%
Decision Tree dengan smote	80%	20%	73%	74%	73%	74%	62%
	70%	30%	78%	76%	76%	76%	72%
	50%	50%	80%	81%	79%	81%	71%

Jika dilihat hasil dari Tabel 7 ketiga model klasifikasi untuk sentiment analisis, terlihat bahwa hasil dari model Support Vector Machine (SVM) dengan menunjukkan hasil dengan tingkat akurasi paling tinggi dengan menggunakan Teknik smote sebesar 93 % dengan evaluasi AUC sebesar 93 %.

Tabel 7. Hasil Keseluruhan

Hasil Keseluruhan Metode							
Metode	Data *Training*	Data *Testing*	*Precision*	*Recall*	F1 *Score*	*Accuracy*	AUC
Naïve Bayes	80%	20%	86%	86%	86%	86%	95%
	70%	30%	85%	85%	84%	85%	89%
	50%	50%	82%	83%	82%	83%	90%
Naïve Bayes dengan *Smote*	80%	20%	90%	89%	89%	89%	95%
	70%	30%	87%	86%	87%	86%	92%
	50%	50%	87%	85%	86%	85%	93%
SVM	80%	20%	90%	89%	87%	89%	89%
	70%	30%	83%	80%	76%	80%	89%
	50%	50%	85%	83%	79%	83%	89%
SVM dengan *Smote*	80%	20%	93%	93%	92%	93%	97%
	70%	30%	85%	85%	84%	85%	86%
	50%	50%	86%	86%	85%	86%	91%
Random Forest	80%	20%	82%	74%	69%	74%	80%
	70%	30%	77%	82%	77%	82%	91%
	50%	50%	75%	78%	73%	78%	83%
Random Forest dengan *smote*	80%	20%	82%	81%	79%	81%	84%
	70%	30%	86%	85%	83%	85%	91%
	50%	50%	80%	80%	77%	80%	87%
Decision Tree	80%	20%	74%	73%	71%	73%	67%
	70%	30%	76%	74%	75%	74%	70%
	50%	50%	78%	78%	77%	78%	70%
Decision Tree dengan *smote*	80%	20%	73%	74%	73%	74%	62%
	70%	30%	78%	76%	76%	76%	72%
	50%	50%	80%	81%	79%	81%	71%

4. Kesimpulan

Hasil analisis sentiment pengguna twitter mengenai layanan indihome mendapatkan hasil crawling data sebanyak 3000 data Hasil crawling tidak di pergunakan semua karena ada tahapan pre-processing yang dimana menghilangkan banyak data karena data tidak valid dan tidak terstruktur serta data yang tidak sempurna. Data tersebut di klasifikasikan menjadi 3 yaitu Negative, Positive, dan Netral [20]. Hasil tahapan pre-processing dan labelling data mendapatkan 405 data yang sudah di beri label, dengan hasil nilai 71,1% Negative, 21,1% Positive dan 7,7% Netral Hasil yang diperoleh dapat disimpulan bahwa pengguna twitter banyak mengeluhkan pelayanan yang diberikan oleh indihome. Namun untuk proses penelitian ini data yang diperoleh tidak balance, agar data balance maka penulis menambahkan library pada python ialah Smote. Hasil perbandingan yang diperoleh ialah Support Vector Machine (SVM) akurasi tanpa SMOTE 89% AUC 95%, akurasi dengan smote 93% AUC 93% dengan data training 80% data testing 20%. Naïve Bayes akurasi tanpa SMOTE 86% AUC 93%, akurasi dengan SMOTE 89% AUC 96% dengan data training 80% data testing 20%. Random Forest akurasi tanpa SMOTE 82% AUC 87% akurasi dengan SMOTE 84% AUC 93% dengan

DOI: https://doi.org/10.38204/tematik.v10i2.1578

data training 70% data testing 30%. Decision Tree akurasi tanpa SMOTE 78% AUC 66% akurasi dengan SMOTE 79% AUC 66% dengan data training 50% data testing 50%. Hasil evaluasi dapat disimpulkan bahwa ternyata smote berhasil membuat tingkat accuracy pada keempat metode ini menjadi naik, SVM dengan SMOTE menunjukkan kinerja yang lebih baik dalam mengklasifikasikan sentimen ulasan tentang Indihome di Twitter. Akurasi yang lebih tinggi dan AUC yang cukup baik menunjukkan bahwa SVM adalah pilihan yang cukup baik untuk analisis sentimen dalam kasus ini. Berdasarkan tujuan penelitian ini juga hasil dari penelitian ini dapat digunakan untuk evaluasi dari pihak indihome juga untuk memberikan pelayanan yang lebih baik lagi.

Daftar Rujukan

[1] P. R. Utami, "Analisis Perbandingan Quality Of Service Jaringan Internet Berbasis Wireless Pada Layanan Internet Service Provider (ISP) Indihome Dan First MediA," vol. 25, no. 2, 2020.

[2] S. Velichety, "Quantifying the impacts of online fake news on the equity value of social media platforms – Evidence from Twitter," *Int. J. Inf. Manag.*, 2022.

[3] A. Diana and M. A. P. Kurniawan, "Decision Support System For Selection Of Internet Service Provider (ISP) With Analytical Hierarchy Process (AHP) And Simple Additive Weighting (SAW) Methods," vol. 4, no. 2, 2022.

[4] A. Puspasari *et al.*, "The Effect Of Service Quality Perception And Company Image On Customer Satisfaction And Their Impact On Customer Loyalty Indihome," vol. 03, no. 02, 2022.

[5] S. F. Pane, R. Prastya, A. G. Putrada, N. Alamsyah, and M. N. Fauzan, "Reevaluating Synthesizing Sentiment Analysis on COVID-19 Fake News Detection using Spark Dataframe," in *2022 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2022, pp. 269–274. doi: 10.1109/ICITSI56531.2022.9970773.

[6] S. H. Suarsa, A. D. Anggraeni, and R. F. Aritonang, "IndiHome Customer Loyalty in Bandung: Service Quality and Price," *Bus. Manag. Res.*, vol. 220.

[7] D. R. Kusuma, S. Syofian, and L. N. Afifa, "Analisis Sentimen Tanggapan Pelanggan Indihome Di Platform Sosial Media Facebook Dan Twitter Menggunakan Support Vector Mesin Dan Pendekatan Klasifikasi Naïve BayeS (Studi Kasus: PT. Telkom Indonesia)," no. 1, 2023.

[8] A. G. Putrada, N. Alamsyah, S. F. Pane, and M. Nurkamal Fauzan, "Feature Importance on Text Analysis for a Novel Indonesian Movie Recommender System," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, 2023, pp. 34–39. doi: 10.1109/ICoICT58202.2023.10262504.

[9] N. Alamsyah, Saparudin, and A. P. Kurniati, "A Novel Airfare Dataset To Predict Travel Agent Profits Based On Dynamic Pricing," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, 2023, pp. 575–581. doi: 10.1109/ICoICT58202.2023.10262694.

[10] S. Winduwati, "Pemanfaatan Media Sosial Instagram Pada Online Shop @ivoree.id dalam Memasarkan Produk," vol. 5, no. 1, 2021.

[11] S. F. Pane, A. G. Putrada, N. Alamsyah, and M. N. Fauzan, "A PSO-GBR Solution for Association Rule Optimization on Supermarket Sales," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 2022, pp. 1–6. doi: 10.1109/ICIC56845.2022.10007001.

[12] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Twitter Social Media Sentiment Analysis Against Bitcoin Cryptocurrency Trends Using Rapidminer," *J. Tek. Inform. Jutif*, vol. 3, no. 5, pp. 1183–1187, Oct. 2022, doi: 10.20884/1.jutif.2022.3.5.289.

[13] A. S. Alhassun and M. A. Rassam, "A Combined Text-Based and Metadata-Based Deep-Learning Framework for the Detection of Spam Accounts on the Social Media Platform Twitter," *Processes*, vol. 10, no. 3, p. 439, Feb. 2022, doi: 10.3390/pr10030439.

[14] A. Halim and Andri Safuwan, "Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization," *J. Inform. Teknol. Dan Sains*, vol. 5, no. 1, pp. 170–178, Feb. 2023, doi: 10.51401/jinteks.v5i1.2229.

[15] R. N. Aziza, T. S. Ardanti, E. Yosrita, and R. F. Ningrum, "Pembangunan Aplikasi Chatbot Informasi Akademik berbasis Cosine Similarity dan Library Sastrawi Stemmer (Studi Kasus: Teknik Informatika IT PLN)," vol. 3, 2022.

[16] S. Akuma, T. Lubem, and I. T. Adom, "Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets," *Int. J. Inf. Technol.*, vol. 14, no. 7, pp. 3629–3635, Dec. 2022, doi: 10.1007/s41870-022-01096-4.

[17] T. Pronk, D. Molenaar, R. W. Wiers, and J. Murre, "Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment," *Psychon. Bull. Rev.*, vol. 29, no. 1, pp. 44–54, Feb. 2022, doi: 10.3758/s13423-021-01948-3.

[18] R. Fahlapi *et al.*, "Analisa Sentimen Vaksinasi Covid-19 Dengan Metode Support Vector Machine Dan Naïve Bayes Berbasis Teknik SmotE," *J. Inform. Kaputama JIK*, vol. 6, no. 1, pp. 57–63, Jan. 2022, doi: 10.59697/jik.v6i1.136.

[19] S. S. Bagui, D. Mink, S. C. Bagui, and S. Subramaniam, "Determining Resampling Ratios Using BSMOTE and SVM-SMOTE for Identifying Rare Attacks in Imbalanced Cybersecurity Data," *Computers*, vol. 12, no. 10, p. 204, Oct. 2023, doi: 10.3390/computers12100204.

[20] D. A. Vonega, A. Fadila, and D. E. Kurniawan, "Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam Pilpres 2024," *J. Appl. Inform. Comput.*, vol. 6, no. 2, pp. 129–135, Nov. 2022, doi: 10.30871/jaic.v6i2.4300.

DOI: https://doi.org/10.38204/tematik.v10i2.1578