

Analisa Sentimen Terhadap Twitter *IndihomeCare* Menggunakan Perbandingan Algoritma *Smote*, *Support Vector Machine*, *AdaBoost* dan *Particle Swarm Optimization*

Ferdian Syah¹, Hanif Fajrin², Abiyyu Nur Afif³, Muhamad Rafi Saeputra^{4*}, Dinda Mirranty⁵, Dedi Dwi Saputra⁶

^{1,2,3,4,5,6} Fakultas Sistem Informasi, Universitas Nusa Mandiri.

article info

Article history:

Received 21 June 2022

Received in revised form

29 October 2022

Accepted 12 November 2022

Available online January 2023

DOI:

<https://doi.org/10.35870/jti.k.v7i1.686>

Keywords:

Indihome; Twitter; Support Vector Machine (SVM); Adaboost; Particle Swarm Optimization (PSO).

abstract

Indihome is one of the largest internet service providers in Indonesia with an increasing number of subscribers every year. Indihome subscribers until the end of March 2022 were recorded at 8.7 million, growing 7.2 percent over the same period last year. However, over time, many customers have complained to IndiHome about slow internet access, sudden increases in billing and so on. Based on the description above, it is interesting to conduct research on the use of Indihome tweets which are the result of channeling opinions and comments on Twitter social media. This study uses the Smote method, Support vector machine, Adaboost and Particle swarm optimization so that the results can be compared with the level of accuracy. The results of this research show that by using the Smote method, Support vector machine obtained values of Accuracy 80.48, Precision 85.29, Recall 73.75 and AUC 0.907. As for the Smote, Support vector machine and AdaBoost methods, the Accuracy values are 80.21, Precision 85.01, Recall 73.36 and AUC 0.861. Finally, the results of the Smote and Particle swarm optimization methods obtained Accuracy values of 76.59, Precision 76.57, Recall 80.35 and AUC 0.868. Based on the research results, the Smote method and the Support vector machine (SVM) have the largest results and are considered effective with the existing dataset.

abstrak

Indihome merupakan salah satu dari sekian internet service provider terbesar di Indonesia dengan peningkatan pelanggan setiap tahunnya. Pelanggan Indihome hingga akhir Maret 2022 tercatat mencapai 8,7 juta, tumbuh 7,2 persen dibanding periode yang sama tahun lalu. Namun, seiring berjalannya waktu, banyak pelanggan yang melayangkan keluhan kepada IndiHome mengenai lambatnya akses internet, kenaikan billing secara tiba-tiba dan lain sebagainya. Berdasarkan deskripsi di atas maka menarik untuk melakukan penelitian penggunaan Indihome tweet yang merupakan hasil dari penyaluran opini dan komentar di sosial media twitter. Penelitian ini menggunakan metode Smote, Support vector machine, Adaboost dan Particle swarm optimization agar hasil tersebut dapat dibandingkan tingkat keakurasiannya. Hasil dari penelitian ini didapat bahwa dengan menggunakan metode Smote, Support vector machine didapat nilai Accuracy 80.48, Precision 85.29, Recall 73.75 dan AUC 0.907. Sedangkan untuk metode Smote, Support vector machine dan AdaBoost didapat nilai Accuracy 80.21, Precision 85.01, Recall 73.36 dan AUC 0.861. Terakhir, hasil dari metode metode Smote dan Particle swarm optimization didapat nilai Accuracy 76.59, Precision 76.57, Recall 80.35 dan AUC 0.868. Berdasarkan hasil penelitian, metode Smote dan Support vector machine (SVM) mempunyai hasil terbesar dan dinilai efektif dengan dataset yang ada.

Kata Kunci:

Indihome; Twitter; Support Vector Machine (SVM); Adaboost; Particle Swarm Optimization (PSO).

* Author. Email: 11212957@nusamandiri.ac.id¹, 11213099@nusamandiri.ac.id², 11213073@nusamandiri.ac.id³, 11212991@nusamandiri.ac.id^{4*}, 11212938@nusamandiri.ac.id⁵, dedi.eis@nusamandiri.ac.id⁶.

1. Latar Belakang

Indihome merupakan salah satu dari sekian *internet service provider* terbesar di Indonesia dengan peningkatan pelanggan setiap tahunnya. Pelanggan Indihome hingga akhir Maret 2022 tercatat mencapai 8,7 juta, tumbuh 7,2 persen dibanding periode yang sama tahun lalu [1]. Namun, seiring berjalannya waktu pula, banyak pelanggan yang melayangkan keluhan kepada Indihome mengenai lambatnya akses internet, kenaikan *billing* secara tiba-tiba dan lain sebagainya. Cara berkomunikasi, seperti adanya keluhan, pengaduan serta saran dari berbagai konsumen dapat dilakukan melalui daring atau media sosial seperti, Twitter yaitu @IndiHomeCare yang dikelola langsung oleh *public relation indihome* [2]. Oleh karena itu, pentingnya seorang *Public Relation Indihome* dalam menjalin hubungan baik antara publik eksternal untuk menjaga keharmonisan dengan publik internal merupakan kewajiban yang harus dilakukan. Berdasarkan deskripsi di atas maka menarik untuk melakukan penelitian penggunaan *indihome tweets* yang merupakan hasil dari penyaluran opini dan komentar, merupakan *resource* yang dapat digunakan untuk menganalisis sentimen khalayak umum terhadap suatu instansi maupun perorangan *twitter* dalam berinteraksi langsung dengan konsumennya.

Mengacu ke penelitian sebelumnya, telah dilakukan penelitian untuk mengukur sentimen analisis dengan berbagai algoritma, diantaranya algoritma *Naïve Bayes* (NB), *Random Forest*, *K-Nearest Neighbour* (K-NN), *Decision Tree*, *K-Medoid*, dan *Backpropagation Neural Network* (BNN) [3][4][5][6][7][8]. Beberapa penelitian menambahkan *feature selection* seperti *Particle Swarm Optimization* (PSO). Diantara penelitian tersebut terdapat beberapa peneliti yang membandingkan dua atau lebih algoritma untuk mendapatkan hasil algoritma terbaik dari algoritma yang dibandingkan. *Particle Swarm Optimization* (PSO) adalah metode optimasi global yang diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995 berdasarkan penelitian terhadap perilaku kawanan burung dan ikan. Setiap partikel PSO memiliki kecepatan partikel dalam ruang pencarian dengan kecepatan yang dinamis disesuaikan dengan perilaku historis mereka [5]. *Smote* merupakan prinsip *oversampling* yaitu menambah data dari kelas minor agar jumlahnya seimbang dengan data kelas mayor [3]. *Adaboost* adalah algoritma yang

ide dasarnya adalah untuk memilih dan menggabungkan sekelompok pengklasifikasi lemah untuk membentuk klasifikasi yang kuat. *Adaboost* adalah algoritma yang iteratif menghasilkan pengklasifikasi dan kemudian menggabungkan mereka untuk membangun klasifikasi utama.[9].

Tweet adalah istilah tulisan atau teks pada layanan *twitter*. Sering kali *twitter* digunakan bukan hanya untuk berbagi informasi, tetapi juga digunakan untuk bercerita tentang sesuatu hal. Tidak jarang juga Ketika bercerita orang-orang melampiaskan emosi. Bercerita dalam *tweet* ini bisa dianalisis dan inilah yang akan menjadi target dalam penulisan ini. *RapidMiner* adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*) dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. *RapidMiner* dikembangkan dengan model *open core* [10].

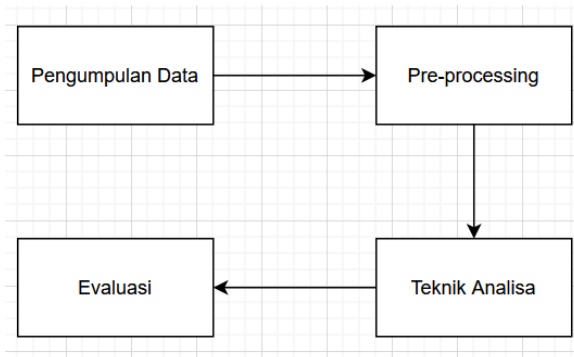
Beberapa penelitian sebelumnya telah membandingkan beberapa algoritma klasifikasi, seperti membandingkan algoritma SVM dengan algoritma NB yang dilakukan pada penelitian dengan judul Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naive Bayes Classifier Dan Support Vector Machine, yang membandingkan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* kemudian mendapatkan hasil akurasi bahwa algoritma *Support Vector Machine* lebih baik dalam menganalisa sentimen daripada algoritma *Naïve Bayes Classifier* dengan tingkat akurasi sebesar 73.96% sedangkan untuk *Naïve Bayes Classifier* adalah 71.94% [11].

Oleh karena itu dapat diformulasikan judul penelitian dengan analisis sentiment ini yaitu Analisa Sentimen Terhadap *Twitter* IndihomeCare Menggunakan Perbandingan Algoritma *Smote*, *Support Vector Machine*, *AdaBoost* Dan *Particle Swarm Optimization*. Secara efektif analisis sentimen termasuk ke dalam salah satu

bidang *Natural Language Processing* yang merupakan cabang ilmu komputer yang mengkaji interaksi antara komputer dengan bahasa alami manusia. Natural Language Processing (NLP) adalah salah satu bidang ilmu computer yang merupakan cabang dari kecerdasan buatan, dan bahasa (linguistik) yang berkaitan dengan interaksi antara komputer dan bahasa alami manusia, seperti bahasa Indonesia atau bahasa Inggris. Tujuan utama dari studi NLP adalah membuat mesin yang mampu mengerti dan memahami makna Bahasa manusia lalu memberikan respon yang sesuai [12]).

2. Metode Penelitian

Pada bagian ini akan dijelaskan mengenai perancangan sistem dalam penelitian ini. Analisis sentimen atau biasa disebut juga dengan *opinion mining* adalah sebuah proses menemukan pendapat pengguna tentang beberapa topik atau teks yang disampaikan pengguna [13]. Dalam pengertian lain, sebuah proses untuk menentukan apakah sepotong tulisan itu bermakna positif, negatif atau netral. Penelitian ini membahas tentang bagaimana melakukan analisis sentimen terhadap data opini pelanggan Indihome pada *twitter* yang digunakan untuk menentukan apakah data opini tersebut masuk opini positif atau negatif.



Gambar 1. Alur sistem

Pengumpulan Data

Pengumpulan data opini masyarakat terhadap Indihome yang dilakukan dengan proses *crawling* dari media *twitter* menggunakan aplikasi *RapidMiner*. Dari data *tweet* yang sudah dikumpulkan sejumlah 1000 data, kemudian dari 1000 data dikelompokkan menjadi 2 terdiri dari *tweet* positif dan *tweet* negatif. Pengelompokan *tweet* dilakukan menggunakan

gataframework dengan memakai 5 teknik, *@Anotation Removal*, *Transformation : Remove URL*, *Regexp*, *Indonesian Stemming* dan *Indonesian Stop word removal*.

Pre-processing

Format:
 Coloumn 1 : No
 Coloumn 2 : Text
 Coloumn 3 : Status

Contoh :

NO	Text	Status
1	Saya suka kamu	Positive
2	Saya tidak suka kamu alias benci	Negative

Masukkan file Excell (.XLS)
 Browse... No file selected.

Techniques 1 : @Anotation Removal
 Techniques 2 : Transformation: Remove URL
 Techniques 3 : Regexp
 Techniques 4 : Indonesian Stemming
 Techniques 5 : Indonesian Stop word removal

Gambar 1. Tahap *Pre-processing* tahap 1

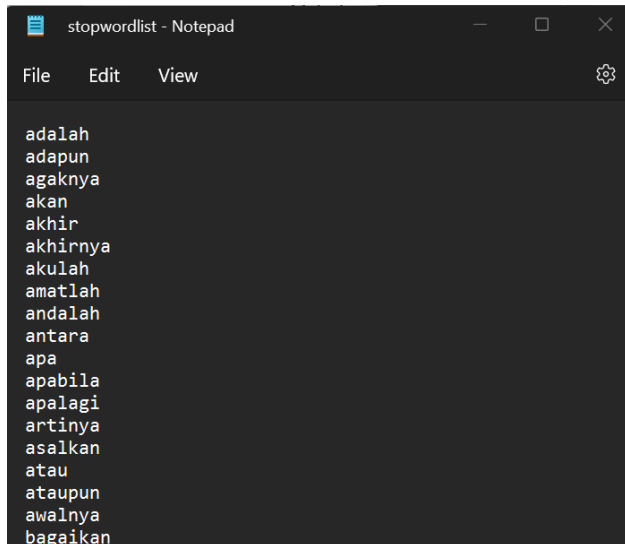
Pre-processing tahap 1 dilakukan menggunakan *website gataframework*. Dimulai dengan *Upload* data excel 2003 yang berisi *dataset* yang berjumlah 1000 data, kemudian penggunaan 5 teknik yaitu *@Anotation Removal*, *Transformation : Remove URL*, *Regexp*, *Indonesian Stemming* dan *Indonesian Stop word removal* dan kemudian data bisa diunduh. Kemudian untuk *pre-processing* tahap 2 dilakukan menggunakan aplikasi *Rapidminer* dengan melakukan proses *Stemming* yaitu merupakan salah satu cara yang digunakan untuk meningkatkan performa IR (*Information Retrieval*) dengan cara menghilangkan imbuhan seperti awalan dan akhiran [14].

```

apa:apaansi
apa:apakh
apa:apasih
benar:benerinnya
aktif:beraktifitas
langgan:berlangganan
langgan:berlangga
langgan:berlanggaan
langgan:berlangganana
langgan:berlanggananan
langgan:langgana
main:bermain
buat:buatin
  
```

Gambar 1. Contoh File *Stemming*.

Setelah proses *stemming* selesai, akan keluar hasil kata-kata yang sudah diperbaiki. Proses selanjutnya yaitu *Stop word removal* yang berguna untuk menghilangkan kata-kata yang tidak dibutuhkan atau tidak berhubungan dengan *dataset* [14].



Gambar 2. Contoh list *stopword*

Teknik Analisa

Kegiatan mengklasifikasikan *tweet* atau analisis sentimen pada penelitian ini dilakukan dengan beberapa metode, yaitu *Smote*, *Support Vector Machine*, *AdaBoost* Dan *Particle Swarm Optimization*. *Support Vector Machine* (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur berdimensi tinggi, digunakan dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning* bias yang berasal dari teori pembelajaran statistik [15].

Evaluasi

Evaluasi dalam penilitan ini digunakan untuk mendapatkan hasil dari kinerja atau performa sebuah metode yang digunakan. Metode yang digunakan pada penulisan ini adalah *Smote*, *Support Vector Machine*, *AdaBoost* dan *Particle Swarm Optimization*. Untuk mendapatkan hasil yang maksimal, perlu dilakukan pengujian dan analisa metode yang tepat dengan dataset yang digunakan.

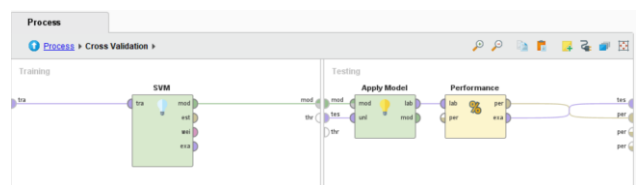
3. Hasil dan Pembahasan

Dataset

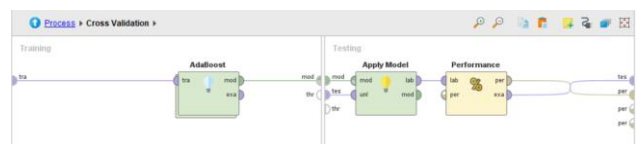
Penggunaan *Dataset* pada penilitan ini adalah dataset yang berasal dari twitter berupa *tweet* dari akun *IndihomeCare* pada tahun 2022. Pengumpulan data dilakukan menggunakan *connection to twitter* melalui aplikasi *Rapid Miner*. Dataset yang dikumpulkan berjumlah 1000 *record* dan kemudian dilakukan proses labelisasi menggunakan *stem* dan *stop word Indonesia removal*. Dari proses pelabelan, didapatkan hasil bahwa *tweet* yang termasuk ke dalam *class* positif sebanyak 653 *tweet*, dan *class* negatif sebanyak 570 *tweet*.

Hasil

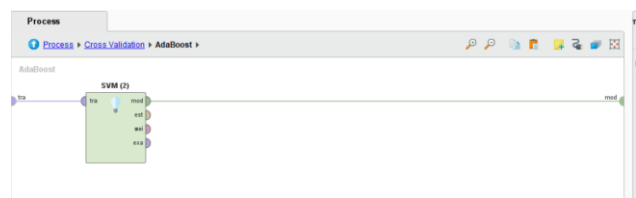
Pada bagian hasil penulis akan menjelaskan hasil pengujian terhadap model usulan yang telah dibuat. Pengujian dilakukan sesuai dengan tahap yang sudah dijelaskan pada bagian metode penelitian. Setelah menjalani proses *pre-processing*, selanjutnya dilakukan analisa perbandingan menggunakan algoritma *Smote*, *Support Vector Machine*, *AdaBoost* dan *Particle Swarm Optimization* sehingga bentuk operator pada aplikasi *Rapidminer* menjadi seperti gambar dibawah ini.



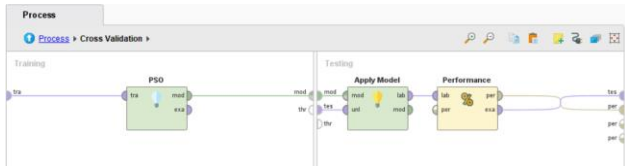
Gambar 3. Penggunaan metode *Smote* dan *Support Vector Machine*



Gambar 4. Penggunaan metode *Smote*, *Support vector machine* dan *AdaBoost*.



Gambar 5. Penggunaan metode *Smote*, *Support vector machine* dan *AdaBoost*.



Gambar 6. Penggunaan metode *Smote* dan *Particle Swarm Optimization*

accuracy: 80.48% +/- 1.77% (micro average: 80.48%)

	true Positive	true Negative	class precision
pred. Positive	648	195	76.87%
pred. Negative	95	548	85.23%
class recall	87.21%	73.76%	

Gambar 7. Hasil *accuracy* menggunakan metode *Smote* dan *Support vector machine*

Pada penggunaan metode *Smote* dan *Support vector machine*, nilai akurasi berada di angka 80.48% seperti pada gambar diatas.

accuracy: 80.21% +/- 4.37% (micro average: 80.22%)

	true Positive	true Negative	class precision
pred. Positive	647	198	76.57%
pred. Negative	96	545	85.02%
class recall	87.08%	73.35%	

Gambar 8. Hasil *accuracy* menggunakan metode *Smote*, *Support vector machine* dan *AdaBoost*.

Sedangkan pada percobaan menggunakan metode *Smote*, *Support vector machine* dan *AdaBoost*, tingkat akurasi yang didapat berada di angka 80.21% seperti gambar diatas.

accuracy: 76.59% +/- 4.88% (micro average: 76.58%)

	true Positive	true Negative	class precision
pred. Positive	541	146	78.75%
pred. Negative	202	597	74.72%
class recall	72.81%	80.35%	

Gambar 9. Hasil *accuracy* menggunakan metode *Smote* dan *Particle Swarm Optimization*.

Terakhir pada percobaan menggunakan metode *Smote* dan *Particle Swarm Optimization* didapat tingkat akurasi pada angka 76.59% seperti gambar diatas. Dari hasil tersebut menunjukan bahwa metode *Smote* dan *Support vector machine* memiliki hasil yang lebih unggul dibandingkan beberapa metode yang lainnya.

4. Kesimpulan

Metode *Smote* dan *Support vector machine* (SVM) mempunyai hasil terbesar dan dinilai efektif dengan dataset yang ada dibanding dengan beberapa metode lainnya. Tujuan adanya penelitian ini yaitu dapat mempermudah analisa terhadap sentimen yang ada di

sosial media *twitter*. Diharapkan untuk penulisan selanjutnya dapat menambahkan metode yang akan digunakan. Penggunaan metode selain *Smote*, *Support Vector Machine*, *AdaBoost* dan *Partial Swarm Optimization* diharapkan mendapatkan hasil yang lebih akurat dari penelitian ini. Hal ini dapat menjadi rujukan dalam penelitian berikutnya dengan menggunakan sumber data yang berbeda.

5. Daftar Pustaka

- [1] Dewi. 2022. Telkom (TLKM) Raup Pendapatan Rp35,20 Triliun, Pelanggan IndiHome Tembus 8,7 Juta - Market Bisnis. [Online] Bisnis.com.
- [2] Mulyono. 2021. Identifikasi Chatbot dalam Meningkatkan Pelayanan Online Menggunakan Metode Natural Language Processing. *Jurnal Informatika Ekonomi Bisnis*, 3(4), 142-147. DOI: <https://doi.org/10.37034/infec.v3i4.102>.
- [3] Syukron, M., Santoso, R. and Widiyarih, T., 2020. Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data. *Jurnal Gaussian*, 9(3), pp.227-236. DOI: <https://doi.org/10.14710/j.gauss.9.3.227-236>.
- [4] Assuja, M.A. and Saniati, S., 2016. Analisis Sentimen Tweet Menggunakan Backpropagation Neural Network. *Jurnal Teknoinfo*, 10(2), pp.48-53. DOI: <https://doi.org/10.33365/jti.v10i2.20>.
- [5] Muhamad, H., Prasojo, C.A., Sugianto, N.A., Surtiningsih, L., Cholissodin, I. and Ilmu, F., 2017. Optimasi naïve bayes classifier dengan menggunakan particle swarm optimization pada data iris. *J. Teknol. Inf. dan Ilmu Komput*, 4(3), p.180.
- [6] Salim, S.S. and Mayary, J., 2020. Analisis Sentimen pengguna Twitter terhadap dompet elektronik dengan metode lexicon based dan k-nearest neighbor. *Jurnal Ilmiah Informatika Komputer*, 25(1), pp.1-17. DOI: <http://dx.doi.org/10.35760/ik.2020.v25i1.2411>.

- [7] Alita, D. and Isnain, A.R., 2020. Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), pp.50-58. DOI: <http://dx.doi.org/10.23960%2Fkomputasi.v8i2.2615>.
- [8] Pradany, L.N. and Fatichah, C., 2016. Analisa Sentimen Kebijakan Pemerintah Pada Konten Twitter Berbahasa Indonesia Menggunakan Svm Dan K-Medoid Clustering. *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, 11(1), pp.59-66.
- [9] Utami, L.D. and Wahono, R.S., 2015. Integrasi metode information gain untuk seleksi fitur dan adaboost untuk mengurangi bias pada analisis sentimen review restoran menggunakan algoritma naive bayes. *Journal of Intelligent Systems*, 1(2), pp.120-126.
- [10] Fadilah, E., 2018. Implementasi Metode Profile Matching Terhadap Sistem Pendukung Keputusan Penerimaan Dana Zakat pada Badan Amil Zakat Pertamina (BAZMA). *Jurnal MATICS Vol*, 10(2). DOI: <https://doi.org/10.18860/mat.v10i2.5745>.
- [11] Rusdaman, D. and Rosiyadi, D., 2019. Analisa Sentimen terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), pp.230-235. DOI: <https://doi.org/10.24114/cess.v4i2.13796>.
- [12] Yunefri, Y., Fadrial, Y.E. and Sutejo, S., 2021. Chatbot Pada Smart Cooperative Oriented Problem Menggunakan Natural Language Processing dan Naive Bayes Classifier. *INTECOMS: Journal of Information Technology and Computer Science*, 4(2), pp.131-140. DOI: <https://doi.org/10.31539/intecom.v4i2.2704>.
- [13] Muhammad, F., 2022. *Analisa Sentimen Mengenai Pendidikan Tatap Muka pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine* (Doctoral dissertation, Universitas Mercu Buana Jakarta).
- [14] Herwijayanti, B., Ratnawati, D.E. and Muflikhah, L., 2018. Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. e-ISSN, pp.306-312.
- [15] Munawarah, R., Soesanto, O. and Faisal, M.R., 2016. Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis. *KLJK-KUMPULAN JURNAL ILMU KOMPUTER*, 3(1), pp.103-113. DOI: <http://dx.doi.org/10.20527/klik.v3i1.39>.