

KLASIFIKASI EMAIL SPAM MENGGUNAKAN ALGORITMA ARTIFICIAL NEURAL NETWORK DAN SUPPORT VECTOR MACHINE

Given Putra

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: 1given.535220143@stu.untar.ac.id

ABSTRAK

Electronic mail (email) merupakan salah satu alat komunikasi yang paling sering digunakan saat ini. Tetapi, kepopuleran email juga membawa masalah, salah satunya adalah maraknya email spam. Email spam bukan hanya mengganggu, tetapi juga bisa membawa dampak buruk bagi seseorang. Maka dari itu, penelitian ini dilakukan dengan tujuan untuk memberikan solusi terhadap permasalahan tersebut, yaitu model machine learning yang dapat mengklasifikasikan email spam. Penelitian ini menerapkan *supervised learning* dengan menggunakan algoritma *artificial neural network* (ANN) serta *support vector machine* (SVM), dan menggunakan data email spam sebagai percobaan. Perbandingan kedua algoritma tersebut dilakukan dengan mengobservasi metrik evaluasi yang dihasilkan dari percobaan. Hasil yang diperoleh dari percobaan adalah algoritma ANN memiliki kinerja yang lebih baik dibandingkan algoritma SVM, yang dapat dilihat dari perbandingan metrik evaluasi kedua algoritma, di mana algoritma ANN mempunyai nilai yang lebih tinggi di segala aspek. Maka, algoritma ANN dapat menjadi algoritma yang baik untuk membangun model yang dapat mendeteksi kehadiran email spam.

Kata kunci—Email, Spam, ANN, SVM, Machine Learning

ABSTRACT

Electronic mail (email) is one of the most frequently used communication tools today. However, the popularity of email also brings problems, one of which is the rise of spam email. Spam emails are not only annoying but can also have a bad impact on someone. Therefore, this research was carried out with the aim of providing a solution to this problem, namely a machine learning model that can classify spam emails. This research applies supervised learning using artificial neural network (ANN) and support vector machine (SVM) algorithms and uses spam email data as an experiment. The comparison of the two algorithms is carried out by observing the evaluation metrics resulting from the experiments. The results obtained from the experiment are that the ANN algorithm has better performance than the SVM algorithm, which can be seen from the comparison of the evaluation metrics of the two algorithms, where the ANN algorithm has a higher score in all aspects. So, the ANN algorithm can be a good algorithm for building a model that can detect the presence of spam emails.

Keywords— Email, Spam, ANN, SVM, Machine Learning

1. PENDAHULUAN

Di era penggunaan *internet* dan media sosial ini, komunikasi *online* telah menjadi aspek penting dalam kehidupan sehari-hari manusia. *Electronic mail (email)* merupakan salah satu sarana komunikasi *online* resmi dan komersial yang paling populer dikarenakan kemudahan mengaksesnya, cepat, dan dapat diandalkan dengan baik. Seiring berjalannya waktu, penggunaan *email* menjadi semakin populer di kalangan masyarakat. Hal tersebut memicu banyaknya *email*

spam yang bermunculan. *Spam* merupakan pesan yang datang dengan motif tersembunyi, seperti menyebarkan iklan, materi promosi lainnya, atau bahkan motif negatif [1]. *Spam* membawa dampak negatif terhadap perangkat yang digunakan, *bandwidth* komunikasi, dan waktu pengguna. Banyak kasus penipuan yang dilakukan oleh pelaku *spam* yang menyamar sebagai bisnis yang sah dalam *email* yang mereka kirimkan dalam upaya mengelabui penerima untuk mengungkapkan informasi pribadinya [2]. Data global terbaru menunjukkan bahwa *email spam* tengah menjadi permasalahan utama bagi para penyedia layanan *email*. Pada bulan Maret 2020, *email spam* telah menyumbang sekitar 53,95% dari lalu lintas dunia [3].

Melihat persoalan tersebut, maka perlu ditemukan solusi untuk mengatasi maraknya *email spam* yang beredar. Beberapa cara telah digunakan untuk membantu mengurangi keberadaan *email spam*, seperti memblokir *internet protocol (IP) address* pengirim atau *email filtering*. Tetapi, seiring berjalannya waktu, *email spam* menjadi semakin canggih dan dapat menghindari cara-cara tersebut [3]. Maka dari itu, penelitian ini mempunyai tujuan sebagai eksperimen untuk mengurangi dan membantu mengatasi *email spam* menggunakan metode *machine learning*. Dari penelitian ini, diharapkan dapat memberikan manfaat berupa model yang bisa mengklasifikasikan antara *email spam* dan tidak melalui parameter dan aturan tertentu. Selain itu, diharapkan bahwa penelitian ini bisa menjadi evaluasi untuk kinerja kedua algoritma yang akan digunakan. Terakhir, diharapkan bahwa penelitian ini bisa menjadi referensi bagi peneliti lainnya di masa depan yang ingin melakukan penelitian di bidang yang sama. Implementasi *machine learning* yang digunakan dalam penelitian ini adalah *supervised learning*, menggunakan dua algoritma, yaitu *artificial neural network (ANN)* dan *support vector machine (SVM)*.

2. METODE PENELITIAN

Pembuatan model klasifikasi *email spam* ini menggunakan dua algoritma implementasi *supervised learning* pada *machine learning*, yaitu *artificial neural network (ANN)* dan *support vector machine (SVM)*.

2.1 Artificial Neural Network (ANN)

Algoritma ANN merupakan algoritma yang dibangun dan dilatih untuk memiliki kemampuan tiruan seperti manusia melalui jaringan *neuron*. *Neuron* merupakan prosesor sederhana yang berjumlah banyak dan saling berhubungan serta bertukar sinyal. Pada ANN, terdapat tiga lapisan jaringan *neuron* yang terdiri dari :

2.1.1 Input Layer

Input layer merupakan lapisan *neuron* yang berfungsi untuk meneruskan data dari luar menuju lapisan berikutnya.

2.1.2 Hidden Layer

Hidden layer merupakan lapisan *neuron* yang berfungsi untuk menghubungkan antara *input* dengan *output* menggunakan hubungan model yang kompleks.

2.1.3 Output Layer

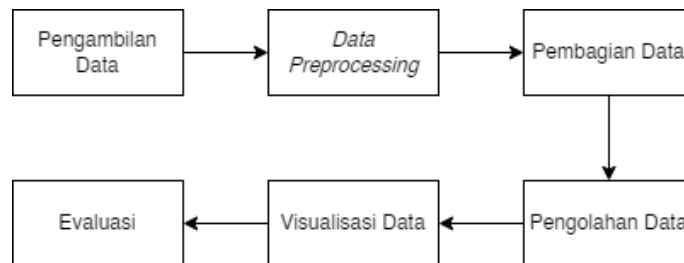
Output layer merupakan lapisan *neuron* yang menjadi tempat keluaran final setelah data

2.2 Support Vector Machine (SVM)

Algoritma SVM merupakan algoritma yang berfungsi dengan cara mencari garis pemisah (*hyperplane*) untuk mendapatkan batas maksimal di antara dua kelas. SVM melakukan klasifikasi dengan memisahkan secara *linear*. Tetapi, SVM juga dapat memisahkan data secara *non-linear* menggunakan fungsi *kernel* seperti *sigmoid*, *radial basis function (RBF)*, atau *polynomial* [6] [7] [8]. SVM mempunyai keunggulan melakukan komputasi data yang memiliki dimensi yang tinggi. Kekurangan SVM terlihat ketika melakukan perhitungan data yang bervolume besar [9].

Dalam prosesnya, penelitian ini melewati enam alur, yaitu pengambilan data, *data*

preprocessing, pembagian data, pengolahan data, visualisasi data, dan evaluasi.



Gambar 1 Alur Penelitian (Sumber : dokumen pribadi)

2.2 Pengambilan Data

Data yang digunakan dalam penelitian ini adalah *dataset email spam*. Data didapatkan melalui situs resmi *Kaggle*, yaitu *www.kaggle.com*. Data terdiri dari 4061 baris (sampel) dan 58 kolom (variabel) dengan tipe data numerik (*integer* dan *float*). Data yang dimaksud tercantum pada gambar 2.

	word_freq_make	word_freq_address	word_freq_all	word_freq_3d	word_freq_our	word_freq_over	word_freq_remove	word_freq_internet	word_freq_order	word_freq_mail
0	0.00	0.64	0.64	0.0	0.32	0.00	0.00	0.00	0.00	0.00
1	0.21	0.28	0.50	0.0	0.14	0.28	0.21	0.07	0.00	0.94
2	0.06	0.00	0.71	0.0	1.23	0.19	0.19	0.12	0.64	0.25
3	0.00	0.00	0.00	0.0	0.63	0.00	0.31	0.63	0.31	0.63
4	0.00	0.00	0.00	0.0	0.63	0.00	0.31	0.63	0.31	0.63

5 rows x 58 columns

	char_freq_;	char_freq(	char_freq[	char_freq!	char_freq\$	char_freq_hash	capital_run_length_average	capital_run_length_longest	capital_run_length_total	spam
0	0.00	0.000	0.0	0.778	0.000	0.000	3.756	61	278	1
0	0.00	0.132	0.0	0.372	0.180	0.048	5.114	101	1028	1
0	0.01	0.143	0.0	0.276	0.184	0.010	9.821	485	2259	1
0	0.00	0.137	0.0	0.137	0.000	0.000	3.537	40	191	1
0	0.00	0.135	0.0	0.135	0.000	0.000	3.537	40	191	1

Gambar 2 Data *Email Spam* (Sumber : *www.colab.research.google.com*)

2.3 Data Preprocessing

Setelah mendapatkan data, langkah selanjutnya adalah memulai *data preprocessing*, yaitu langkah di mana data melakukan pemrosesan awal sebelum diolah. Tujuan dari proses ini adalah menghasilkan data final yang bersih tanpa adanya elemen yang tidak relevan. Pertama, periksa terlebih dahulu apakah terdapat *missing values* pada data. *Missing values* merupakan data yang tidak memiliki nilai pada kolom [10]. Pada gambar 3 terlihat bahwa semua variabel bersifat *non-null*, yang artinya tidak terdapat *missing values* pada data.

#	Column	Non-Null Count	Dtype
0	word_freq_make	4601 non-null	float64
1	word_freq_address	4601 non-null	float64
2	word_freq_all	4601 non-null	float64
3	word_freq_3d	4601 non-null	float64
4	word_freq_our	4601 non-null	float64
5	word_freq_over	4601 non-null	float64
6	word_freq_remove	4601 non-null	float64
7	word_freq_internet	4601 non-null	float64
8	word_freq_order	4601 non-null	float64
9	word_freq_mail	4601 non-null	float64
10	word_freq_receive	4601 non-null	float64
11	word_freq_will	4601 non-null	float64
12	word_freq_people	4601 non-null	float64
13	word_freq_report	4601 non-null	float64
14	word_freq_addresses	4601 non-null	float64
15	word_freq_free	4601 non-null	float64
16	word_freq_business	4601 non-null	float64
17	word_freq_email	4601 non-null	float64
18	word_freq_you	4601 non-null	float64
19	word_freq_credit	4601 non-null	float64
20	word_freq_your	4601 non-null	float64
21	word_freq_font	4601 non-null	float64
22	word_freq_data	4601 non-null	float64
23	word_freq_money	4601 non-null	float64
24	word_freq_hp	4601 non-null	float64
25	word_freq_hpl	4601 non-null	float64
26	word_freq_george	4601 non-null	float64
27	word_freq_550	4601 non-null	float64
28	word_freq_lab	4601 non-null	float64
29	word_freq_labs	4601 non-null	float64
30	word_freq_telnet	4601 non-null	float64
31	word_freq_857	4601 non-null	float64
32	word_freq_data	4601 non-null	float64
33	word_freq_415	4601 non-null	float64
34	word_freq_85	4601 non-null	float64
35	word_freq_technology	4601 non-null	float64
36	word_freq_1999	4601 non-null	float64
37	word_freq_parts	4601 non-null	float64
38	word_freq_pm	4601 non-null	float64
39	word_freq_direct	4601 non-null	float64
40	word_freq_cs	4601 non-null	float64
41	word_freq_meeting	4601 non-null	float64
42	word_freq_original	4601 non-null	float64
43	word_freq_project	4601 non-null	float64
44	word_freq_re	4601 non-null	float64
45	word_freq_edu	4601 non-null	float64
46	word_freq_table	4601 non-null	float64
47	word_freq_conference	4601 non-null	float64
48	char_freq_;	4601 non-null	float64
49	char_freq(	4601 non-null	float64
50	char_freq[	4601 non-null	float64
51	char_freq!	4601 non-null	float64
52	char_freq\$	4601 non-null	float64
53	char_freq_hash	4601 non-null	float64
54	capital_run_length_average	4601 non-null	float64
55	capital_run_length_longest	4601 non-null	int64
56	capital_run_length_total	4601 non-null	int64
57	spam	4601 non-null	int64

dtypes: float64(55), int64(3)

Gambar 3 Tidak Terdapat *Missing Values* (Sumber : *www.colab.research.google.com*)

2.4 Pembagian Data

Tahap selanjutnya adalah menentukan skenario pembagian data untuk proses pengolahan. Pada penelitian ini, data dibagi menjadi dua, yaitu data uji dan data latih [11]. Skenario pembagian data pada penelitian ini dibagi menjadi tiga, yaitu 10% data uji dan 90% data latih, 20% data uji dan 80% data latih, serta 30% data uji dan 70% data latih.

2.5 Pengolahan Data

Pengolahan data dilakukan menggunakan implementasi dari algoritma ANN dan SVM yang sudah dijelaskan. *Random state* yang digunakan untuk pengolahan data adalah 0, 10, 20, 30, dan 40.

2.6 Visualisasi Data

Untuk memudahkan proses penelitian, maka hasil dari pengolahan data dari kedua algoritma tersebut dituangkan dalam bentuk tabel disertai dengan deskripsi mengenai tabel tersebut.

2.7 Evaluasi

Hasil dari visualisasi data dapat digunakan untuk mengkaji keseluruhan dari penelitian dan melihat bagaimana performa klasifikasi menggunakan kedua algoritma tersebut. Metode evaluasi yang digunakan adalah dengan metrik evaluasi *precision*, *accuracy*, *recall*, *f1-score*. Berikut merupakan penjelasan mengenai masing-masing metrik evaluasi tersebut :

2.7.1 Precision

Precision merupakan metrik jumlah klasifikasi yang sesuai dengan permintaan informasi.

2.7.2 curacy

Accuracy merupakan metrik perbandingan keakuratan yang didapatkan antara nilai prediksi dengan nilai yang sebenarnya.

2.7.3 call

Recall merupakan metrik yang menandakan seberapa banyak *instance* positif yang berhasil ditemukan oleh model dalam pengolahan data.

2.7.4 F1-Score

F1-score merupakan metrik yang menunjukkan keseimbangan antara metrik *precision* dengan metrik *recall* [4] [12].

3. HASIL DAN PEMBAHASAN

Variabel *output* hasil klasifikasi pada data terletak pada variabel “*spam*” (lihat gambar 2), di mana jenis klasifikasi ini merupakan klasifikasi biner atau *binary classification*. Klasifikasi biner mempunyai arti bahwa *output* yang dihasilkan hanya terdapat dua, yang biasanya adalah 1 atau 0 [13]. Dalam data yang digunakan, 1 mengartikan bahwa pesan *email* tersebut adalah *email spam*. Sedangkan 0 mengartikan bahwa pesan *email* tersebut tidak termasuk kategori *email spam* [14]. Hal ini ditentukan melalui parameter dan peraturan yang ditentukan oleh fitur-fitur pada data.

3.1 Pembahasan Hasil Algoritma ANN

Pengolahan data menggunakan algoritma ANN menggunakan tiga skenario dan *random state* 0, 10, 20, 30, dan 40.

3.1.1 Pembahasan Hasil Algoritma ANN dengan 10% Data Uji dan 90% Data Latih

Tabel 1 Hasil Algoritma ANN dengan 10% Data Uji dan 90% Data Latih

Eksperimen ke	Precision	Accuracy	Recall	F1-Score
1	0.93	0.92	0.91	0.92
2	0.96	0.95	0.95	0.95
3	0.93	0.94	0.93	0.94
4	0.89	0.90	0.90	0.90
5	0.93	0.93	0.93	0.93
Rata-rata	0.93	0.93	0.92	0.93

3.1.2 Pembahasan Hasil Algoritma ANN dengan 20% Data Uji dan 80% Data Latih

Tabel 2 Hasil Algoritma ANN dengan 20% Data Uji dan 80% Data Latih

Eksperimen ke	Precision	Accuracy	Recall	F1-Score
1	0.93	0.93	0.92	0.93
2	0.94	0.94	0.94	0.94
3	0.89	0.87	0.85	0.86
4	0.90	0.91	0.90	0.90
5	0.94	0.94	0.93	0.93
Rata-rata	0.92	0.92	0.91	0.91

3.1.3 Pembahasan Hasil Algoritma ANN dengan 30% Data Uji dan 70% Data Latih

Tabel 3 Hasil Algoritma ANN dengan 30% Data Uji dan 70% Data Latih

Eksperimen ke	Precision	Accuracy	Recall	F1-Score
1	0.87	0.88	0.89	0.88
2	0.93	0.93	0.92	0.92
3	0.93	0.93	0.93	0.93
4	0.92	0.92	0.90	0.91
5	0.89	0.87	0.84	0.86
Rata-rata	0.91	0.91	0.90	0.90

3.1.4 Keseluruhan Rata-Rata Percobaan Algoritma ANN

Tabel 4 Keseluruhan Rata-Rata Percobaan Algoritma ANN

Data uji	Data latih	Precision	Accuracy	Recall	F1-Score
10%	90%	0.93	0.93	0.92	0.93
20%	80%	0.92	0.92	0.91	0.91
30%	70%	0.91	0.91	0.90	0.90

3.2 Pembahasan Hasil Algoritma SVM

Pengolahan data menggunakan algoritma *SVM* dengan *kernel RBF* dan nilai *C* adalah 10, menggunakan tiga skenario dan *random state* 0, 10, 20, 30, dan 40.

3.2.1 Pembahasan Hasil Algoritma SVM dengan 10% Data Uji dan 90% Data Latih

Tabel 5 Hasil Algoritma SVM dengan 10% Data Uji dan 90% Data Latih

Eksperimen ke	Precision	Accuracy	Recall	F1-Score
1	0.74	0.72	0.70	0.70
2	0.81	0.77	0.74	0.75
3	0.76	0.75	0.72	0.72
4	0.75	0.75	0.69	0.70
5	0.77	0.77	0.73	0.74
Rata-rata	0.77	0.75	0.72	0.72

3.2.2 Pembahasan Hasil Algoritma SVM dengan 20% Data Uji dan 80% Data Latih

Tabel 6 Hasil Algoritma SVM dengan 20% Data Uji dan 80% Data Latih

Eksperimen ke	Precision	Accuracy	Recall	F1-Score
1	0.77	0.75	0.71	0.72
2	0.78	0.76	0.71	0.72
3	0.76	0.75	0.71	0.71
4	0.73	0.73	0.68	0.69
5	0.73	0.74	0.69	0.70
Rata-rata	0.75	0.75	0.70	0.71

3.2.2 Pembahasan Hasil Algoritma SVM dengan 30% Data Uji dan 70% Data Latih

Tabel 7 Hasil Algoritma SVM dengan 30% Data Uji dan 70% Data Latih

Eksperimen ke	Precision	Accuracy	Recall	F1-Score
1	0.75	0.73	0.69	0.70
2	0.74	0.73	0.69	0.69
3	0.75	0.73	0.69	0.69
4	0.72	0.73	0.68	0.69
5	0.72	0.73	0.68	0.69
Rata-rata	0.74	0.73	0.69	0.69

3.2.3 Keseluruhan Rata-Rata Percobaan Algoritma SVM

Tabel 8 Keseluruhan Rata-Rata Percobaan Algoritma SVM

Data uji	Data latih	Precision	Accuracy	Recall	F1-Score
10%	90%	0.77	0.75	0.72	0.72
20%	80%	0.75	0.75	0.70	0.71
30%	70%	0.74	0.73	0.69	0.69

3.3 Hasil Evaluasi

Berdasarkan percobaan yang telah dilakukan, terlihat pada tabel 4 dan 8 bahwa algoritma ANN memiliki kinerja yang lebih baik dibandingkan algoritma SVM. Hal ini dapat disimpulkan dari rata-rata metrik *precision*, *accuracy*, *recall*, dan *f1-score* yang diperoleh. Semuanya dimenangi oleh algoritma ANN bahkan dengan berbagai data uji dan data latih yang berbeda. Sebagai contoh, rata-rata metrik *accuracy* percobaan algoritma ANN dengan data uji 10% dan data latih 90% bernilai 0.93, yang di mana hasil tersebut lebih tinggi sekitar 25% dari metrik *accuracy* percobaan algoritma SVM dengan data uji 10% dan data latih 90%, yaitu 0.73. Berdasarkan percobaan menggunakan data yang tersedia, algoritma ANN memberikan kinerja yang sangat memuaskan untuk mengklasifikasi dan mendeteksi *email spam*, dengan semua nilai metrik yang sudah menyentuh angka minimal 0.90. Sebaliknya, kinerja algoritma SVM terhadap data tersebut terbilang relatif rendah, hanya berkisaran di 0.70, bahkan ada yang di bawah angka tersebut. Hal ini menunjukkan bahwa perlu diadakan penelitian lebih lanjut mengenai penyebab metrik evaluasi SVM yang rendah serta cara untuk meningkatkan nilai-nilai metrik tersebut untuk membuat model pendeteksi *email spam* yang lebih baik lagi menggunakan algoritma SVM.

4. KESIMPULAN

Dari keseluruhan percobaan yang telah dilakukan, terdapat beberapa kesimpulan yang bisa diambil, yaitu :

4.1 Algoritma ANN Dapat Menjadi Algoritma yang Tepat untuk Mendeteksi Email Spam

Melihat kinerja algoritma ANN yang memuaskan, dapat disimpulkan bahwa untuk membuat model yang mempunyai kemampuan untuk mengklasifikasikan *email spam* dapat menggunakan algoritma ANN.

4.2 Kinerja Algoritma ANN Lebih Baik Daripada Algoritma SVM

Berdasarkan percobaan yang telah dilakukan menggunakan data *email spam*, algoritma ANN memberikan performa yang lebih baik daripada algoritma SVM. Hal ini dapat dilihat dari seluruh metrik evaluasi (*precision*, *accuracy*, *recall*, dan *f1-score*) dari algoritma ANN lebih tinggi daripada metrik evaluasi algoritma SVM.

4.3 Kemungkinan Pengembangan Model

Saran penulis untuk peneliti lainnya yang ingin melakukan penelitian serupa di masa yang akan datang adalah untuk menggunakan data yang lebih luas dan besa *volume* nya, agar model yang diciptakan semakin akurat dan kuat untuk menghadapi *email spam* dalam bentuk apapun. Modifikasi berbagai variabel tertentu pada suatu algoritma merupakan hal yang penting untuk mengoptimalkan kinerja suatu model. Penulis juga menyarankan untuk mengeksplor kemungkinan penggunaan algoritma lain, baik di dalam *supervised learning* maupun di luar

DAFTAR PUSTAKA

- [1] I. AbdulNabi and Q. Yaseen, "Spam Email Detection Using Deep Learning Techniques," *Procedi Computer Science*, vol. 184, pp. 853-858, 2021.
- [2] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi and O. E. Ajibuw: "Machine learning for Email Spam Filtering: Review, Approaches and Open Research Problems," *Heliyon*, vol. 5, no. 6, 2019.
- [3] M. Jazzar, R. F. Yousef and D. Eleyan, "Evaluation of Machine Learning Techniques for Email Spar Classification," *I. J. Education and Management Engineering*, vol. 4, pp. 35-42, 2021.
- [4] D. G. Pradana, M. L. Alghifari, M. F. Juna and S. D. Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, pp. 55-60, 2022.
- [5] A. E. Anthoni, N. A. Abdullah and P. S. S. Palaniappan, "Email Spam Classification Tool Usin Artificial Neural Network (ANN) with Logistic Activation Function," *Applied Informatio Technology and Computer Science*, vol. 1, pp. 15-21, 2020.
- [6] R. A. Rizal, I. S. Girsang and S. A. Prasetyo, "Klasifikasi Wajah Menggunakan Support Vect Machine (SVM)," *REMIK: Riset Dan E-Jurnal Manajemen Informatika Komputer*, vol. 3, pp. 1-2019.
- [7] M. N. Yarkuran, S. and F. Aziz, "Klasifikasi Kategori Obat Menggunakan Algoritma Support Vect Machine," *Journal Pharmacy and Application of Computer Sciences*, vol. 1, pp. 27-32, 2023.
- [8] O. H. Rahman, G. Abdillah and A. Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosi Twitter Menggunakan Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 5, pp. 17-23, 2021.
- [9] P. Arsi, R. Wahyudi and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacar Pindah Ibu Kota Indonesia," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 5, p. 231-237, 2021.
- [10] P. Misra and A. S. Yadav, "Impact of Preprocessing Methods on Healthcare Predictions," *Proceeding of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 201* 2019.
- [11] R. A. Nadir and R. N. Sukmana, "Sistem Prediksi Harga Emas Berdasarkan Data Time Seri Menggunakan Metode Artificial Neural Network (ANN)," *Digital Transformation Technolog (Digitech)*, vol. 3, pp. 426-437, 2023.
- [12] S. A. Pratiwi, A. Fauzi, S. A. P. Lestari and Y. Cahyana, "Prediksi Persediaan Obat Pada Apote Menggunakan Algoritma Decision Tree," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, p. 2381-2388, 2024.
- [13] D. Berrar, "Performance Measures for Binary Classification," pp. 546-560, 2019.
- [14] K. N. A. Halim, A. S. M. Jaya and A. F. A. Fadzil, "Data Pre-Processing Algorithm for Neur Network Binary Classification Model in Bank Tele-Marketing," *International Journal of Innovati Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 3, 2020.