

## KLASIFIKASI PENGGUNAAN DATA TRAFIK INTERNET MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

Pitrasacha Adytia<sup>1</sup>, Wahyuni<sup>2</sup>, Kelik Sussolaikah<sup>3</sup> dan Yudha Satria<sup>4</sup>

<sup>1,2,4</sup>Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia

<sup>3</sup>Program Studi Teknik Informatika, Universitas PGRI Madiun, Madiun, Indonesia

<sup>1</sup>Email: [pitra@wicida.ac.id](mailto:pitra@wicida.ac.id)

<sup>2</sup>Email: [wahyuni@wicida.ac.id](mailto:wahyuni@wicida.ac.id)

<sup>3</sup>Email: [kelik@unipma.ac.id](mailto:kelik@unipma.ac.id)

<sup>4</sup>Email: [yuda.satria024@gmail.com](mailto:yuda.satria024@gmail.com)

### ABSTRAK

Tak dapat dipungkiri saat ini internet merupakan sesuatu yang sangat penting untuk berbagai kebutuhan. Tidak terkecuali di STMIK Widya Cipta Dharma. Internet banyak digunakan dalam lingkungan kampus, baik oleh mahasiswa, dosen dan juga tenaga kependidikan. Kegiatan belajar mengajar dan juga pekerjaan dalam lingkungan kampus tidak terlepas dari kebutuhan penggunaan internet. Namun waktu penggunaan internet juga terkadang menumpuk dalam jam-jam tertentu dan menyebabkan kecepatan internet menjadi lambat. Hal itu dipengaruhi oleh banyaknya pengiriman paket *header* pada *flow*/arus lalu lintas internet sehingga koneksi menjadi berat dan terasa lambat. Oleh karena itu, dibutuhkan suatu metode klasifikasi yang dapat memberikan informasi mengenai aktivitas mahasiswa, dosen dan tenaga kependidikan dalam penggunaan internet. Adapun algoritma klasifikasi yang digunakan adalah klasifikasi *Support Vector Machine* (SVM). Metode pengembangan yang digunakan adalah SKKNI Nomor 299 Tahun 2020. Parameter yang digunakan adalah arus paket yang dikirim oleh *user* dan paket yang diterima oleh *user*. Adapun hasil penelitian ini berupa model algoritma SVM yang dapat mengklasifikasikan arus penggunaan trafik internet dengan empat kategori yaitu Download, Game, SocialNetwork, dan Web yang memiliki akurasi 64% dengan menggunakan kernel *Radial Basis Function* (RBF). Hasil akurasi yang dihasilkan cukup rendah dan membuat algoritma SVM tidak cocok untuk melakukan klasifikasi trafik internet dan perlunya metode lain untuk mengklasifikasikan trafik internet.

Kata kunci: internet, klasifikasi, *support vector machine*

### ABSTRACT

It is undeniable that nowadays, the internet is essential for various needs. STMIK Widya Cipta Dharma is no exception. The internet is widely used in the campus environment by students, lecturers and education staff. Teaching and learning activities and work in the campus environment are inseparable from the need to use the internet. However, internet usage time sometimes accumulates in certain hours and causes slow internet speeds. This is influenced by the large number of header packets sent in the internet traffic flow, so the connection becomes heavy and feels sluggish. Therefore, a classification method is needed to provide information about the activities of students, lecturers and academic staff using the internet. The classification algorithm used is the Support Vector Machine (SVM). The development method used is SKKNI Number 299 of 2020. The parameters used are the flow of packets sent by the user and packets received by the user. The results of this study are in the form of an SVM algorithm model that can classify current internet traffic usage into four categories, namely Download, Game, SocialNetwork, and Web, which has an accuracy of 64% using the Radial Basis Function (RBF) kernel. The resulting accuracy results are pretty low and make the SVM algorithm unsuitable for classifying internet traffic and the need for other methods to classify internet traffic.

Keywords: internet, classification, support vector machine

### 1. PENDAHULUAN

Tak dapat dipungkiri saat ini internet merupakan sesuatu yang sangat penting untuk berbagai kebutuhan. Tidak terkecuali di STMIK Widya Cipta Dharma. Internet banyak digunakan dalam lingkungan kampus, baik oleh mahasiswa, dosen dan juga tenaga kependidikan. Kegiatan belajar mengajar dan juga pekerjaan dalam lingkungan kampus tidak terlepas dari kebutuhan penggunaan internet. Namun waktu penggunaan internet juga terkadang menumpuk dalam jam-jam tertentu. Dan biasanya di jam-jam tersebut kecepatan internet menjadi lambat. Hal itu dipengaruhi oleh banyaknya pengiriman paket header pada *flow*/arus lalu lintas internet yang sehingga koneksi menjadi berat.

Mengklasifikasikan lalu lintas jaringan dan menghubungkan lalu lintas jaringan dengan aplikasi yang dihasilkan adalah merupakan langkah penting pertama untuk menganalisis jaringan, informasi berharga dapat dikumpulkan dari analisis trafik, terutama untuk tujuan keamanan seperti menyaring data

trafik dan mengidentifikasi serta mendeteksi aktivitas jahat [1]. Dengan mengetahui jenis trafik apa yang mengalir melalui jaringan, maka operator jaringan dapat bereaksi dengan cepat terhadap potensi insiden yang tidak diinginkan.

Untuk melakukan klasifikasi secara aktual berdasarkan karakteristik statistik, pengklasifikasi perlu menggunakan teknik mining, khususnya algoritma machine learning, karena perlu menangani pola trafik yang berbeda. Algoritma machine learning sangat ringan dan secara komputasi lebih murah daripada teknik klasifikasi berbasis payload, karena tidak bergantung pada *Deep Packet Inspection* (DPI) melainkan memanfaatkan informasi dari analisis tingkat arus/flow, keefektifan pengklasifikasi dalam klasifikasi statistik bergantung pada fitur yang diekstraksi dari arus, yang membutuhkan pengetahuan luas karena kompleksitasnya, namun teknik ini mengungguli teknik berbasis payload karena tidak menangani isi paket, dan demikian dapat menganalisis arus trafik terenkripsi tanpa kesulitan [2].

Setiap tahun, metode pengenkripsian data semakin dikembangkan oleh para ilmuwan dan ahli kriptografi. Karena metode enkripsi data yang sudah cukup tua sangat rentan terhadap pembobolan data. Beberapa metode atau algoritma enkripsi yang sering digunakan dalam pengenkripsian data diantaranya adalah: Rivest, Shamir and Adleman (RSA), Base64 dan lain sebagainya. Setelah mengetahui pembahasan tentang metode- metode klasifikasi trafik pada jaringan internet. Maka dalam penulisan ini peneliti akan menerapkan metode klasifikasi SVM untuk membuat model yang dapat mengklasifikasikan trafik jaringan internet, sehingga dari klasifikasi tersebut dapat diketahui jenis trafiknya.

Adapun ruang lingkup dalam penelitian ini permasalahan mencakup, penerapan Support Vector Machine pada klasifikasi data trafik internet berbasis arus paket, data yang digunakan adalah dari hasil menangkap trafik internet di Pusat Komunikasi (Puskom) STMIK Wicida Widya Cipta Dharma Samarinda yang berbentuk .pcap yang didapat dari proses menangkap langsung dengan perangkat lunak wireshark kemudian dilakukan proses labelling dengan NStream, model dapat mengklasifikasikan 4 kelas yaitu Download, Game, SocialNetwork dan Web dengan menggunakan data latih sebanyak 6000 data, atribut yang digunakan dalam fase modeling sebanyak 11 atribut arus paket trafik internet.

## 2. MATERI DAN METODE

### Pengembangan data Mining SKKNI

Beberapa tahapan dalam melakukan proses klasifikasi menggunakan tahapan pengembangan data mining berdasarkan SKKNI Nomor 299 Tahun 2020 dapat dilihat pada tabel 1.

Tabel 1. Pengembangan data Mining SKKNI

Fungsi Utama	Fungsi Dasar
Bussiness Understanding	Menentukan objektif bisnis Menentukan tujuan teknis data science Membuat rencana proyek
Data understanding	Mengumpulkan data Menelaah data Memvalidasi data
Data preparation	Memilih data Membersihkan data Mengkonstruksi data Menentukan label Mengintegrasikan data
Modeling	Membangun skenario pengujian Membangun model
Model evaluation	Mengevaluasi hasil pemodelan Melakukan review proses
Deployment	Membuat rencana deployment model Melakukan deployment model Membuat rencana pemeliharaan Melakukan pemeliharaan model
Evaluation	Melakukan review proyek data science Membuat laporan akhir proyek data science

### Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah salah satu metode klasifikasi yang berdasarkan pada diskriminan linear margin maksimum. Adapun tujuan dari diskriminan linear margin maksimum adalah untuk mencari *hyperplane* dengan memaksimalkan jarak atau margin antar kelas [3]. Cara terbaik untuk mencari *hyperplane* terbaik antara kedua kelas adalah dengan mengukur margin *hyperplane* dan kemudian mencari titik maksimalnya. Margin merupakan jarak antara *hyperplane* dengan data terdekat dari masing-

masing kelas. Data yang paling dekat ini disebut sebagai *support vector* [4]. Adapun langkah-langkah dalam SVM sebagai berikut:

Menentukan nilai  $w$  dan  $b$  dengan menggunakan persamaan 1.

$$y_i(w \cdot x_i + b) \geq 1, i = 1,2,3, \dots, N \dots\dots\dots (1)$$

di mana  $w$  adalah vector bobot,  $x$  nilai masukan atribut data,  $b$  nilai bias dan  $y$  adalah label ke- $i$ .

Masukkan data uji kedalam persamaan dengan syarat jika hasil persamaan bernilai positif maka diklasifikasikan sebagai kelas positif (persamaan 2):

$$w \cdot x_i + b > 1 \dots\dots\dots (2)$$

Jika hasil persamaan bernilai negative maka diklasifikasikan sebagai kelas negatif (persamaan 3):

$$w \cdot x_i + b \leq -1 \dots\dots\dots (3)$$

**Multi Class SVM**

Pada saat pertama kali dipernalkan pada tahun 1992 oleh Vapnik, SVM hanya dapat mengklasifikasikan data ke dalam dua kelas. SVM dikembangkan menjadi *multiclass* yang artinya dapat mengklasifikasikan kelas menjadi lebih dari dua kelas. Ada dua pendekatan dalam menggunakan SVM *multiclass*, salah satunya adalah dengan menggabungkan beberapa SVM biner [5]. Metode penggabungan beberapa SVM biner adalah sebagai berikut:

1. *One-Against-All*

Metode ini menggunakan prinsip satu lawan semua dengan membandingkan satu kelas dengan semua kelas lainnya. Ketika akan melakukan klasifikasi data ke dalam  $k$  kelas, maka harus dibangun pula sejumlah  $k$  model SVM biner. Setiap model biner SVM ke- $i$  akan dilatih dengan menggunakan keseluruhan data agar ditemukan apakah merupakan bagian dari kelas ke- $i$  atau bukan ketika diklasifikasikan.

2. *One-Against-One*

Metode *one against one* atau satu lawan satu ini akan membandingkan satu kelas dengan kelas lainnya dalam membangun sejumlah model SVM. Ketika akan melakukan klasifikasi data ke dalam  $k$  kelas, maka diharuskan untuk membangun sejumlah model dengan persamaan 4:

$$k(k - 1) / 2 \dots\dots\dots (4)$$

**K-Fold Cross Validation**

*Cross validation* adalah sebuah metode dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi maksimum ketika data dibagi menjadi dua subset (data latih dan data uji) [6]. *K-fold cross validation* adalah metode yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai  $K$ . Pada pendekatan metode *K-fold cross validation*, dataset dibagi menjadi sejumlah buah partisi secara acak. Data partisi tersebut diolah sejumlah  $K$  kali eksperimen dengan masing-masing eksperimen menggunakan data partisi ke- $K$  sebagai data *testing* dan menggunakan sisa partisi lainnya sebagai data *training* [7].

**Normalisasi Min-Max**

Normalisasi Min-Max merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli. Proses normalisasi bertujuan untuk memetakan nilai dari masing-masing variabel ke dalam rentang yang sama yakni rentang [0,1], sehingga pada saat proses perhitungan nilai *similarity*, masing-masing variabel memberikan tingkat kepentingan yang sama (memberikan pengaruh yang sama) [8]. Normalisasi ini dapat dilakukan dengan menggunakan persamaan 5.

$$x' = \frac{x - \text{nilai}_{\min}}{\text{nilai}_{\max} - \text{nilai}_{\min}} \dots\dots\dots (5)$$

**Confusion Matrix**

*Confusion matrix* adalah salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi [9]. Pada pengukuran kinerja menggunakan *confusion matrix* (tabel 2), terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai TN merupakan jumlah data negatif yang terdeteksi dengan benar. Nilai FP merupakan data negatif namun terdeteksi sebagai data positif. TP merupakan data positif yang terdeteksi benar. FN merupakan kebalikan dari TP, di mana data positif namun terdeteksi sebagai data negatif.

Tabel 2. *Confusion matrix*

		Kelas hasil prediksi		
		Ya	Tidak	Jumlah
Kelas	Ya	TP	FN	P
	Tidak	FP	TN	N
	Jumlah	P	N	P + N

### Library Scikit-Learn

Scikit-Learn atau yang biasa dikenal dengan sklearn adalah sebuah perangkat lunak *open-source* yang biasa digunakan untuk *machine learning* dalam bahasa pemrograman Python [10], *library* ini mendukung berbagai macam algoritma seperti klasifikasi, regresi, dan *clustering*, termasuk algoritma *k-means*, Scikit-Learn juga merupakan salah satu *library* yang paling populer di Github.

### NFStream

NFStream adalah sebuah framework Python yang menyediakan struktur data yang cepat, fleksibel, dan ekspresif yang dirancang untuk dapat bekerja dengan data trafik jaringan *online* atau *offline* dengan mudah [11]. Kemampuan utama dari NFStream adalah dapat menghitung statistik *flow* yang di-*generate* dari informasi trafik yang diambil di Wireshark seperti fitur *flow* (jumlah total *byte* dari semua paket milik dari *flow* tertentu), fitur *flow* berasal dari *header* paket IP, TCP, dan UDP. Perhitungannya melibatkan metode analisis-sentris, seperti menentukan ringkasan statistik (minimum maksimum, *mean* dan standar deviasi) dari panjang paket dan waktu antar-kedatangan. Selain menghitung statistik *flow*, NFStream juga dapat menentukan jenis aplikasi yang didapat dari meng-*generate flow*, fitur ini dicapai dengan mengintegrasikannya dengan *library* nDPI.

### Tahapan Penelitian

Adapun tahapan penelitian adalah sebagai berikut:

1. *Bussiness Understanding*, pada tahapan ini proses yang dilakukan adalah melihat masalah yang terjadi pada STMIK Widya Cipta Dharma mengenai trafik internet.
2. *Data Understanding*, pada tahapan ini dilakukan pengumpulan data dan memahami isi data tersebut. Pengumpulan data dilakukan dengan meng-*capture* data trafik. Data yang diambil adalah data yang didapatkan melalui hasil *capture* wifi selama lima hari di STMIK Widya Cipta Dharma. Setelah itu dilakukan tahapan *generate* data trafik. Setelah mendapatkan data trafik dengan format *file* (.pcap), maka setelahnya dilakukan proses *generate* data menggunakan NFStream. Proses *generate* akan menghasilkan sebuah file Comma Separator Value (CSV) yang akan dilakukan proses *data mining*.
3. *Data Preparation*, setelah melakukan analisis terhadap data yang sudah ada, selanjutnya dilakukan proses persiapan data dengan data trafik internet Puskom WiCiDa yang memiliki 45 atribut. Namun data tersebut tidak dapat langsung digunakan. Data tersebut harus dibersihkan terlebih dahulu sebelum digunakan untuk tahapan *modelling*.
4. *Modeling*, pada tahap ini dilakukan proses dengan algoritma klasifikasi yang dipilih menghasilkan suatu pola informasi yang dapat memudahkan peneliti. Pola klasifikasi yang dihasilkan dengan atribut yang sudah dipilih pada tahap sebelumnya akan dapat digunakan untuk melakukan proses klasifikasi kategori trafik internet Puskom dengan menggunakan sejumlah atribut yang mempengaruhinya, pada tahap ini dilakukan proses pembangunan model yang menerapkan pembelajaran SVM di dalamnya.
5. *Model Evaluation*, setelah model telah melakukan training dan mendapatkan pola, selanjutnya dilakukan tahap *model evaluation* dengan teknik evaluasi *confusion matrix*, yang dapat memberikan tampilan informasi perbandingan hasil kinerja dari algoritma terhadap data uji. *Confusion matrix* digunakan untuk mengukur performa dalam permasalahan klasifikasi biner ataupun permasalahan *multiclass* dengan melihat nilai *accuracy*, *precision*, dan *recall*.
6. Pengujian, pada tahap ini dilakukan proses pengujian dengan menggunakan *K-fold cross validation* yang akan melakukan proses *training* dan pengujian setiap *fold* untuk mendapatkan gambaran terhadap kualitas model dari setiap model yang diuji. Peneliti menguji dengan menerapkan kernel RBF, Polynomial, dan linear pada model SVM dengan menghasilkan rata-rata *output* hasil perbandingan dalam bentuk akurasi.

## 3. HASIL DAN PEMBAHASAN

Pada tahapan *bussiness understanding*, setelah dianalisis dan diperhatikan arus data trafik internet di unit Pusat Komputer (PUSKOM) STMIK Widya Cipta Dharma selama lima hari mulai pukul 09.00 – 16.00, dapat disimpulkan bahwa dalam rentang waktu tersebut ada waktu-waktu yang memiliki trafik internet yang sangat padat, sehingga membuat akses internet menjadi lambat. Sehingga akan dilakukan klasifikasi menggunakan SVM *multiclass* untuk mengklasifikasikan arus paket trafik internet. Dan diklasifikasi ke dalam empat kategori, yaitu Download, Game, SocialNetwork, Web. Dan akan dilakukan analisis penggunaan kernel yang memiliki akurasi paling tinggi.

Pada tahapan *data understanding*, data trafik internet yang sudah diolah menggunakan NFStream mengandung informasi mengenai paket-paket data yang sudah dilakukan proses perhitungan statistika. Paket-paket data ini mendeskripsikan jumlah paket data yang dikirim oleh user dalam mengakses internet. Adapun atribut dan kelas yang digunakan dalam proses penelitian ini, dapat dilihat pada tabel 3.

Tabel 3. Deskripsi fitur trafik internet

Tipe Atribut	Deskripsi Atribut	Jumlah
Paket Src2Dst	Arus paket total, byte, mean, max, min dan standar deviasi yang dikirim dari alamat asal ke alamat tujuan	6
Paket Dst2Src	Arus paket total, byte, mean, max, min dan standar deviasi yang dikirim dari alamat tujuan ke alamat asal	6
Bidirectional Paket	Arus paket total, byte, mean, max, min dan standar deviasi yang dikirim dari dua arah	6
Paket Interval Src2Dst	Waktu antar paket min, mean, max dan standar deviasi yang dikirim dari alamat asal ke alamat tujuan	4
Paket Interval Dst2Src	Waktu antar paket min, mean, max dan standar deviasi yang dikirim dari alamat tujuan ke alamat asal	4
Bidirectional Interval	Waktu antar paket min, mean, max dan standar deviasi yang dikirim dari dua arah	4
Src Alamat	Alamat asal yang terdiri dari (ip, mac, oui, dan port)	4
Dst Alamat	Alamat tujuan yang terdiri dari (ip, mac, oui, dan port)	4
Protocol	Nomor identifikasi transport layer protokol	1
Application name	Nama aplikasi	1
Application_category_name	Kategori arus data trafik	1
Src_dst_address_and_src_dst_port	Nomor unik alamat asal, port dan tujuan, port	1
Time_Hours	Waktu jam aktivitas arus data trafik	1
Day	Waktu hari aktivitas arus data trafik	1
Day_Time	Waktu hari dan jam aktivitas arus data trafik	1

Terdapat 27 kategori yang dihasilkan dari pengambilan data selama lima hari. Kategori yang diakses terbanyak adalah kategori Web sebanyak 503.131 data. Adapun kategori yang paling sedikit diakses adalah kategori IOT Scada, sebanyak 17 data.



Gambar 1. Data kategori trafik selama lima hari

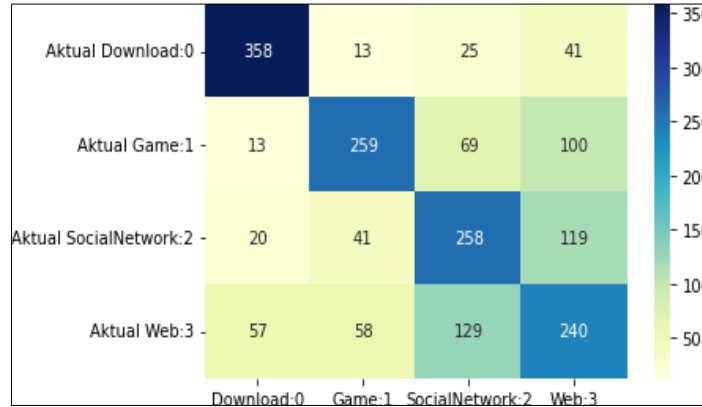
Selanjutnya ke tahapan *Data Preparation* dilakukan proses pemilihan atribut sebanyak 11 fitur. Adapun data yang digunakan adalah sebanyak 6000 data yang mana jumlah data tersebut adalah hasil data trafik internet yang didistribusikan berdasarkan empat kategori yaitu Web, Download, SocialNetwork dan Game. Dari 6000 data disebar untuk masing-masing kategori dengan masing-masing jumlah data pada tiap kategori adalah 1500. Hal tersebut dilakukan agar terjadi *balancing* data.

*Modelling*, pada tahapan ini data akan dipisah berdasarkan data latih sebanyak 4200 dan data uji sebanyak 1800, kemudian peneliti melakukan proses normalisasi MinMax pada data latih dan data uji serta pada label dengan LabelEncoder. Peneliti menggunakan algoritma SVM untuk permasalahan klasifikasi multiclass di mana Scikit-Learn menyediakan sebuah model yang bernama SVC(), yang akan membuat sebuah model baru dengan menggunakan pembelajaran algoritma SVM, peneliti menggunakan fungsi ini dan memilih kernel RBF untuk proses training, karena RBF adalah kernel terbaik yang dapat memisahkan empat buah *class* dan memiliki nilai akurasi tertinggi dibandingkan kernel linear dan *polynomial* ini dikarenakan kedua parameter tersebut cenderung memiliki proses optimisasi pada parameter tidak cocok dalam mengklasifikasi data trafik internet.

Model *evaluation*, peneliti menggunakan visual *confusion matrix* sebagai cara untuk menunjukkan di mana model membuat prediksi yang benar dan membuat prediksi yang salah. Evaluation dapat dilihat



pada Gambar 4 di mana model dapat memprediksi dengan benar sebanyak 358 jenis trafik Download dan error dengan total 90 kesalahan dari himpunan jenis trafik [Game, SocialNetwork, Web], lalu model dapat memprediksi dengan benar jenis trafik Game sebanyak 254 data namun dengan salah prediksi jumlah data sebanyak 112 dari total kelas [Download, SocialNetwork, Web], lalu kelas SocialNetwork memprediksi dengan benar sebanyak 258 data dengan kesalahan total prediksi dari himpunan [Download, Game, Web] sebesar 223 dan kelas Web dapat diprediksi dengan benar sebanyak 240 namun gagal memprediksi data yang lain dengan total himpunan jenis trafik [Download, Game, SocialNetwork] sebesar 260 data berdasarkan hasil prediksi benar pada setiap kelas, ternyata model cenderung dapat memprediksi kelas Download lebih baik dari kelas lainnya.



Gambar 2. Confusion matrix

Pengujian, dilakukan untuk melihat apakah kinerja model dengan kernel RBF dapat bekerja dengan baik. Proses pengujian akan dibagi menjadi dua skenario yaitu pengujian dengan menggunakan kernel dan fitur scaling Min-Max, dan pengujian tanpa fitur scaling.

Tabel 4. Hasil Pengujian Akurasi Kernel Dengan MinMax

Fold	Akurasi Kernel		
	Linear	RBF	Polynomial
1	0,42	0,63	0,35
2	0,43	0,64	0,39
3	0,41	0,63	0,37
4	0,44	0,66	0,39
Rata- rata	0,42	0,64	0,37

Berdasarkan tabel 4, rata-rata akurasi kernel yang digunakan adalah sebagai berikut:

1. Kernel Linear memiliki rata-rata sebesar 42%.
2. Kernel RBF memiliki rata-rata sebesar 64%.
3. Kernel Polynomial memiliki rata-rata sebesar 37%.

Dari hasil yang telah dijabarkan, maka dapat dilihat bahwa kernel yang paling baik digunakan dengan menggunakan MinMax adalah kernel RBF.

Tabel 5. Hasil Pengujian Akurasi Kernel Tanpa MinMax

Fold	Akurasi Kernel		
	Linear	RBF	Polynomial
1	0,378	0,248	0,248
2	0,417	0,237	0,237
3	0,416	0,24	0,243
4	0,402	0,22	0,234
Rata- rata	0,40	0,24	0,24

Adapun hasil akurasi kernel tanpa menggunakan MinMax berdasarkan tabel 5 adalah sebagai berikut:

1. Kernel Linear memiliki rata-rata sebesar 40%.
2. Kernel RBF memiliki rata-rata sebesar 24%.
3. Kernel Polynomial memiliki rata-rata sebesar 24%.

Dari hasil penjabaran sebelumnya, rata-rata akurasi kernel tanpa MinMax tidak lebih baik dibandingkan dengan akurasi kernel menggunakan MinMax. Rata-rata dari semua kernel adalah kurang dari 50%.

#### 4. KESIMPULAN DAN SARAN

Kesimpulan yang bisa diperoleh dari penelitian ini telah menghasilkan sebuah model SVM yang dapat mengklasifikasikan empat kategori trafik internet yaitu Download, Game, SocialNetwork dan Web dengan kernel RBF dan normalisasi MinMax mendapatkan akurasi sebesar 62%, yang artinya algoritma

SVM kurang bagus dalam mengklasifikasikan trafik internet dengan menggunakan arus/flow paket dua arah dan data yang dinormalisasi dapat mempengaruhi hasil akurasi dari sistem klasifikasi. Pada skenario pengujian RBF Kernel dengan *K-fold cross validation* mendapatkan hasil akurasi sebesar 64%. Namun mendapatkan penurunan yang tajam saat pengujian tanpa normalisasi data dan mendapatkan akurasi rendah yaitu 24%, yang artinya tanpa menggunakan normalisasi MinMax data yang diuji akan mendapatkan akurasi yang rendah. Penelitian ini dapat dikembangkan dengan membuat model yang menggunakan atribut arus satu arah dan menggunakan teknik *hyperparameter tuning* untuk mencari parameter yang membuat model mendapatkan akurasi yang lebih tinggi.

#### DAFTAR PUSTAKA

- [1] S. Riyadi, "Penerapan Metode Naive Bayes dalam Pengklasifikasi Trafik Jaringan," *SMATIKA Jurnal*, vol. 6, no. 2, p. 29, 2016, doi: [10.32664/smatika.v6i02.45](https://doi.org/10.32664/smatika.v6i02.45)
- [2] N. Al Khater and R. E. Overill, "Network traffic classification techniques and challenges," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*, Oct. 2015, doi: [10.1109/ICDIM.2015.7381869](https://doi.org/10.1109/ICDIM.2015.7381869).
- [3] M. J. Zaki and J. W. Meira, *Data Mining and Analysis*. Cambridge University Press, 2014. doi: [10.1017/CBO9780511810114](https://doi.org/10.1017/CBO9780511810114).
- [4] E. Prasetyo, *Data Mining : mengolah data menjadi informasi menggunakan Matlab*. Yogyakarta: Andi, 2014.
- [5] Suyanto, *Data mining : untuk klasifikasi dan klusterisasi data*. Bandung: Penerbit Informatika, 2017.
- [6] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, May 2022, doi: [10.14710/j.gauss.v11i1.33991](https://doi.org/10.14710/j.gauss.v11i1.33991).
- [7] T. Kurniawan, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media *Mainstream* Menggunakan *Naïve Bayes Classifier* dan *Support Vector Machine*," Skripsi Institut Teknologi Sepuluh Nopember, Surabaya, 2017.
- [8] M. Nishom, "Perbandingan Akurasi *Euclidean Distance*, *Minkowski Distance*, dan *Manhattan Distance* pada Algoritma *K-Means Clustering* berbasis *Chi-Square*," *Jurnal Pengembangan IT (JPIT)*, Vol. 4, No. 1, pp. 20–24, Jan. 2019, doi: [10.30591/jpit.v4i1.1253](https://doi.org/10.30591/jpit.v4i1.1253).
- [9] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163–170, September 2019, doi: [10.33480/pilar.v15i2.658](https://doi.org/10.33480/pilar.v15i2.658).
- [10] N. A. Rakhmawati, D. B. Aletha, L. P. G. Widiastuti, A. S. Laka Kaki, and A. T. Aminullah, "Klusterisasi Peraturan Daerah di Seluruh Wilayah Jawa dengan Menggunakan Algoritma *K-means*," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 2, pp. 242–246, Juli 2020, doi: [10.32736/sisfokom.v9i2.816](https://doi.org/10.32736/sisfokom.v9i2.816).
- [11] Z. Aouini and A. Pekar, "NFStream," *Computer Networks*, vol. 204, Februari 2022, doi: [10.1016/j.comnet.2021.108719](https://doi.org/10.1016/j.comnet.2021.108719).