ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

Visualisasi Trafik Jaringan Dengan Metode *Support Vector Machine (SVM)* (Studi Kasus: Universitas Indo Global Mandiri)

Tasmi[1)], **Reza Maulana**[2)], **Husnawati**[3)]

[1)] *Program Studi Sistem Komputer Universitas Indo Global Mandiri*
[2),3)] *Program Studi Sistem Komputer Universitas Indo Global Mandiri*
Jl. Srijaya Negara Bukit Lama Palembang, Jl. Jendral Sudirman No.629 Km 4 Palembang
Email : 0301622126022@student.unsri.ac.id [1)], reza@uigm.ac.id [2)], husnawati@uigm.ac.id [3)]

ABSTRACT

Limited network resources and the increasing number of internet users in the current digital era have an impact on high traffic which results in decreased access speed to internet services. This is also a problem that occurs at the Indo Global Mandiri University (UIGM) Palembang, causing access to academic services to be slow. The purpose of this research is to identify the types of network traffic patterns which are then carried out by the process of grouping and visualizing these types of traffic. The data in this study were taken in real-time at the UIGM campus. The data obtained is the result of responses which are then extracted. The extraction results are processed using the Support Vector Machine (SVM) method for the process of grouping and visualizing data. The results of this study can distinguish types of traffic based on communication protocols, namely tcp and udp, where the results of the experiment were carried out six times with the results being the first experiment where 99.7% TCP and 0.1% for UDP, the second experiment 97.6% for TCP and 1.1% for UDP , trial three 99.7 % TCP and 0.2% UDP, trial four 97.5% and 1.3% UDP, trial five 99.5 TCP and 02% UDP, and the sixth or final try 97.7% TCP and 1.1% UDP. The data from the use of the SVM method obtained several types of traffic such as games by 0.4%, mail 0.2%, multimedia 0.4% and the web by 82.8% and this research still produces data that the pattern is not yet recognized by 15.5%

Keywords : *Network Traffic, Classification, Support Vector Mesin*

ABSTRAK

Keterbatasan sumber daya jaringan serta semakin meningkatnya pengguna internet pada era digital saat ini berdampak pada tingginya trafik yang mengakibatkan menurunnya kecepatan akses pada layanan internet. Hal ini juga merupakan permasalahan yang terjadi di Universitas Indo Global Mandiri (UIGM) Palembang, sehingga menyebabkan akses kelayanan akademik menjadi lambat. Tujuan dari penelitian ini adalah untuk pengenalan jenis pola trafik jaringan yang selanjutny dilakukan proses pengelompokan dan visualisasi jenis tarfik tersebut. Data pada penelitian ini diambil secara real-time pada kampus UIGM. Data yang diperoleh merupakan hasil tanggapan yang kemudian diekstrak. Hasil ekstraksi tersebut diolah menggunakan metode Support Vector Machine (SVM) untuk proses pengelompokan dan visualisasi data. Hasil penelitian ini dapat membedakan jenis trafik berdasarkan protokol komunikasi yaitu tcp dan udp, dimana hasil percobaan dilakukan sebanyak enam kali dengan hasilnya adalah percobaan pertama dimana 99,7% TCP dan 0.1 % untuk UDP, percobaan kedua 97.6% untuk TCP serta 1.1% untuk UDP, percobaan tiga 99.7 % TCP dan 0.2% UDP, percobaan empat 97,5% dan 1.3 % UDP, percobaan lima 99,5 TCP dan 02% UDP, serta percobaan ke enam atau terakhir didapatkan 97.7% TCP dan 1.1% UDP. Data hasil penggunaan metode SVM didapatkan beberapa jenis trafik seperti games sebesar 0.4%, mail 0,2% multimedia 0,4% dan web sebesar 82,8% serta peneltian ini masih dihasilkan data nya belum dikenali polahnya sebesar 15.5%.

Kata Kunci : *Trafik Jaringan, Klasifikasi, Support Vector Mesin*

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

1. Pendahuluan

Meningkatnya penggunaan internet dan semakin berkembangnya layanan yang tersedia, ditambah terbatasnya sumberdaya jaringan berdampak pada tingginya trafik jaringan yang berakibat pada lambatnya akses ke internet. Pengelolaan sumberdaya jaringan dan analisis trafik data diperlukan dalam rangka mengoptimalkan sumberdaya jaringan dan menyediakan layanan internet yang stabil (Awad et al., 2016),(Liu et al., 2018).

(Shafiq et al., 2017) pada penelitiannya menyatakan dengan mengetahui pola dan mengklasifikasikan trafik jaringan dapat mengoptimalkan sumber daya jaringan dan mengetahui pertumbuhan trafiknya. Hal serupa juga dikemukakan oleh (Yang, Narantuya, & Lim, 2019) klasifikasi trafik jaringan sangatlah penting untuk menganalisa sumber daya jaringan dalam rangka menyediakan layanan jaringan yang aman dan berkualitas. Mengenali pola trafik dan mengklasifikasikannya menjadi solusi yang efektif dalam pengelolaan dan optimalisasi sumberdaya jaringan.

Klasifikasi trafik jaringan berdasarkan *IP Address*, *Port Number* dan *Protocol* sudah tidak efektif lagi hal ini disebabkan oleh modifikasi pola data dan pengirimannya, sebagai contoh seperti penggunaan lebih dari satu alamat *IP Public* untuk satu layanan, penggunaan *Port* dan *Protocol* secara dinamis sehingga sulit untuk mengenali pola trafik secara akurat (Zamfir, Balan, Sandu, & Costache, 2016). Klasifikasi trafik jaringan secara akurat dapat dilakukan salah satunya dengan penerapan *Machine Learning*, informasi yang terdapat pada *Header TCP* seperti *IP Dst*, *Port Dst*, *Protocol*, *Payload*, *Time Stamp* dan *TTL* akan dijadikan fitur untuk mengenali pola trafik yang akan diklasifikasikan (Networks, Dhote, Agrawal, & Deen, 2015).

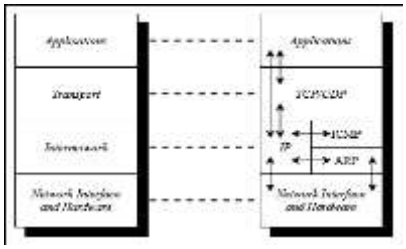
Support Vector Machine (SVM) salah satu metode pada *Machine Learning* yang dapat terapkan untuk klasifikasi tafik jaringan, SVM bekerja berdasarkan pencarian fitur dengan bobot dan parameter terbaik untuk menentukan batas distribusi dan mengelompokkannya. Kelebihan dari metode SVM adalah tingkat akurasi klasifikasi yang tinggi, tetapi untuk menghasilkan akurasi yang tinggi diperlukan sampel data yang besar. Kelemahan dari metode SVM adalah pada waktu proses yang sangat berpengaruh pada besarnya data, semakin besar data maka waktu proses akan semakin lama (Hao et al., 2015).

Tingginya trafik jaringan dan terbatasnya sumberdaya jaringan yang tersedia juga menjadi salah satu permasalah di Universitas Indo Global Mandiri (UIGM), hal ini berdampak pada lambatnya akses internet terutama di jam sibuk. Pengelolaan sumberdaya jaringan selama ini yang hanya mengatur distribusi pembagian *Bandwidth* belum dapat mengatasi masalah tersebut, maka perlu diketahui pola trafik jaringan untuk mendapatkan data aktifitas penggunaan jaringan. Setelah diketahui pola trafik dan data penggunaan jaringan, maka dapat ditentukan prioritasnya berdasarkan kebutuhan yang paling utama. Berdasarkan latar belakang yang telah diuraian, maka pada penelitian ini akan dilakukan visualisasi dan pengelompokan trafik jaringan dengan metode SVM di UIGM

A. Tinjaun Pustaka

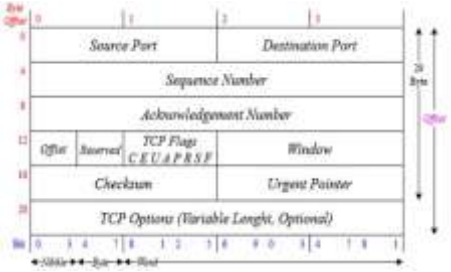
(Robertazzi, 2017) Teknologi jaringan komputer merupakan sebuah kumpulan *node* yang saling terhubung melalui media transmisi yang mampu berkomunikasi satu sama lain. Jaringan komputer dapat dibedakan menjadi beberapa kategori berdasarkan sebagai berikut: (i). Media Transmisi, (ii). Letak Geografis, (iii). Topologi dan (iv) Layanan dan koneksi.

Untuk dapat berkomunikasi dan melakukan pertukaran data dalam sebuah jaringan diperlukan protokol-protokol yang dapat menghubungkan beberapa komputer dengan sistem operasi dan mesin yang berbeda. *TCP/IP* merupakan kumpulan protokol protokol komunikasi yang pertama kali dikembangkan oleh ARPANET dan DARPA.



Gambar 1. Lapisan Protokol TCP/IP (Parziale et al., 2006)

Proses pengiriman informasi atau data dari sumber ke tujuan dibutuhkan identitas atau ciri-ciri dari setiap layer di OSI seperti ditampilkan pada gambar 2 berikut.



Gambar Error! No text of specified style in document.. Struktur TCP Header (Parziale et al., 2006)

1. *Source Port*, memiliki ukuran 2 *byte* (16 *bit*) berfungsi mengindikasikan sumber *Protocol* lapisan aplikasi yang mengirimkan segmen *TCP* bersangkutan. Gabungan antara *Field Source IP Address* dalam *Header IP* dan *Field Source Port* pada *Field Header TCP* disebut juga sebagai *Socket Number*, yang artinya sebuah alamat global dari mana segmen dikirimkan.

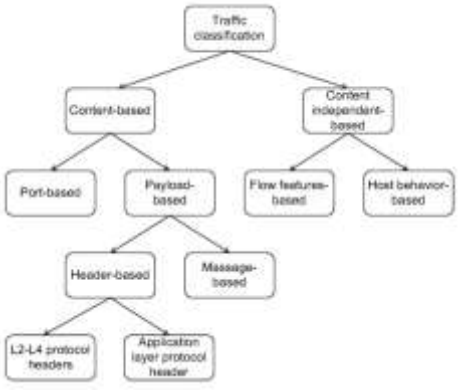
ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

2. *Destination Port*, memiliki ukuran 2 *byte* (16 *bit*) berfungsi mengindikasikan tujuan *Protocol* lapisan aplikasi yang menerima segmen *TCP* bersangkutan, yang artinya sebuah alamat global ke mana segmen akan dikirmkan.
3. *Sequence Number*, berukuran 4 *byte* (32 bit) mengindikasikan nomor urut dari *Octet* pertama dari data di dalam sebuah segmen *TCP* yang akan dikirimkan.
4. *Acknowledgment Number*, berukuran 4 *byte*. *Ack Number* mengindikasikan nomor urut dari *octet* selanjutnya dalam aliran *byte* yang diharapkan diterima oleh pengirim dari sisi penerima pada pengiriman selanjutnya. *ACK* sangat penting pada segmen-segmen *TCP* dengan *Flag ACK* diset ke nilai 1.
5. *Flags* (6 *bit*), mengindikasikan *Flag-Flag TCP* seperti *URG, ACK, PSH, RST, SYN,* dan *FIN*.
6. *Window* (2 *byte*/ 16 *bit*), mengindikasikan jumlah *byte* yang tersedia dimiliki oleh *Buffer Host* penerima segmen bersangkutan. Tujuannya adalah untuk mengatur lalu lintas data atau *Flow Control.*
7. *Checksum* (2 *byte* /16 *bit*), melakukan pengecekan integritas segmen *TCP(Header* dan *Payload)*. Nilai *Field Checksum* akan diatur ke nilai 0 selamaproses kalkulasi *Checksum.*

1. Klasifikasi Trafik Jaringan

Klasifikasi trafik dibagian dalam dua cara yaitu *content- based* dan *content independent-basde. Content Based* adalah metode mengenali pola dari sebuah trafik seperti: port, protokol, IP address (*port-based, Payload-based*), sedangkan *Content Independent-Based* (Siqueira, Ruiz, & Loureiro, 2007)

Klasifikasi trafik jaringan merupakan proses identifikasi aplikasi dan protokol terhadap kondisi pada jaringan, terdapat beberapa teknik klasifikasi seperti penelitian yang dilakukan oleh (Shafiq et al., 2017) melakukan proses klasfikasi menggunakan *Port Based Technique,* tetapi hasilnya kurang akurat untuk aplikasi yang menggunakan *Dynamic Port.*



Gambar 3. Proses klasifikasi trafik (Siqueira et al., 2007)

Port Based Technique, merupakan teknik klasifikasi berdasarkan *Port Number*, klasifikasi dengan teknik ini kurang akurat untuk aplikasi yang menggunakan *Dynamic Port.*

Tabel 1. *Well Known Port* yang terdaftar di IANA

Assigned Ports	***Applications***
20	*FTP Data*
21	*FTP*
22	*SSH*
23	*Telnet*
53	*DNS*
80	*HTTP*
110	*POP3*
123	*NTP*
161	*SNMP*
443	*HTTPS*

Payload-Based Technique, teknik ini disebut juga dengan *DPI.* Pada teknik ini identifikasi trafik dikenali dari karakteristik pola trafik. Tabel 2 terdapat beberapa contoh *Signature* dari beberapa aplikasi yang dapat digunakan untuk mengenali trafik pada jaringan

Tabel Error! No text of specified style in document.**.**
Contoh Signature Beberapa Aplikasi

Applications	***Signature***	***Protocol***
Edonkey 2000	0xe3190100000xe53f010000	*TCP/UDP*
Fasttrack	"Get /.hash"0x2700000002980	*TCP/UDP*
BitTorrent	"0x12Bit"	*TCP*
Gnutella	"GNUT" "GIV"	*TCP*
Aress	"GET hash", "Get Shal"	*UDP*

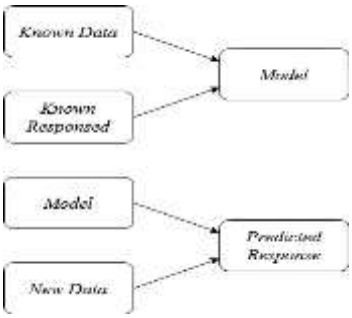
2. *Machine Learning*

Machine Learning atau pembelajaran mesin merupakan salah satu kecerdasan buatan yang memungkinkan mesin melakukan pembelajaran berdasarkan contoh data, pembelajaran mesin memanfaatkan hubungan antar variabel dan probabilitas untuk menghasilkan prediksi (Boutaba et al., 2018).

Salah satu metode yang paling banyak mendapat perhatian dalam bidang pengenalan pola yaitu *support vector machine* disingkat SVM (Byun & Lee, 2003). SVM merupakan pembelajaran yang didasari pada teori statistik yang dikombinasikan dengan teori-teori belajar yang telah ada seperti konsep *kernel trick*, teori umum, metode optimasi, dan lain sebagainya SVM mampu mengklasifikasi data secara *linear* maupun *non-linear*

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

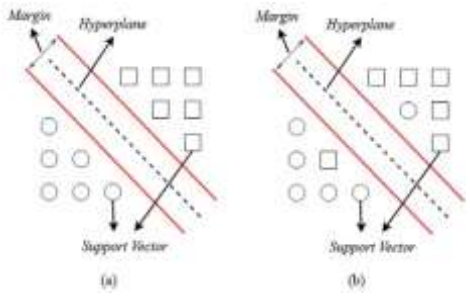
Berdasarkan masukan dan keluaran yang diharapkan, pembelajaran mesin terbagi menjadi dua kelompok yaitu : *Supervised Learning*, merupakan pembelajaran yang bertujuan untuk memetakan masukan dan keluaran yang diinginkan seperti pada pengelompokkan.



Gambar 4. Deskripsi *Supervised Learning* (Shafiq et al., 2017)

Gambar 2 menjelaskan prinsip kerja *Supervised Learning* yaitu dengan mempelajari sekumpulan contoh masukan dan keluaran dan menghasilkan sebuah model yang mampu memetakan masukan yang baru menjadi keluaran yang tepat. Terdapat beberapa algoritma yang dapat diterapkan antara lain, *Naïve Bayes, SVM* dan *Decision Tree. Unsupervised Learning*, merupakan pembelajaran yang memodelkan himpunan masukan untuk mempelajari dan mencari pola-pola tertentu pada masukan yang diberikan, *Clustering* atau penggolongan merupakan penerapan dari pembelajaran ini. Beberapa algoritma pada pembelajaran ini antara lain, *Birch*, *Cure* dan *K-Mean* (Shafiq et al., 2017).

Support Vector Machine (SVM) merupakan salah satu metode yang terdapat pada *Supervised Learning*, metode *SVM* pada penerapannya terdiri dari 2 tahapan yaitu *Training dan Testing.* Metode *SVM* bertujuan untuk membangun sebuah model yang mampu memprediksi data uji ke sebuah data baru dengan mempelajari hubungan antar fitur yang dilambangkan dengan x_i $(x \epsilon X)$ dan penentuan kelas yang dilambangkan dengan y_i $(y \epsilon Y)$, $F = f(x_i) = h_\Theta(x_i)$ merupakan fungsi pemetaan hubungan dimana Θ merupakan bobot matriks dari fungsi pemetaan. Untuk menentukan nilai *F*, harus didapatkan matriks Θ terlebih dahulu.



Gambar 5. (a). *Linear Classification* (b). *Non-Linear Classification* (Kong, Huang, Wu, Tang, & Ye, 2018)

Gambar 3. menjelaskan jika data bersifat linier, *SVM* akan mencari linier *Hyperplane* untuk membagi data menjadi kelas-kelas biner dengan margin maksimum. (b). menjelaskan jika data bersifat tidak linier, maka dibutuhkan perubahan fitur awal ke ruang dimensi yang lebih tinggi. *SVM* menyediakan 4 kernel yaitu : *Linear Kernel, Polynomial Kernel, Radial Basis Function (RBF) Kernel* dan *Sigmoid Kernel* yang dapat digunakan dalam memecahkan permasalahan data yang tidak linier (Kong et al., 2018)

3. Data Visualization

Data Visualization adalah teknik penyajian informasi dalam bentuk visual yang bertujuan untuk memberikan pemahaman dan analisis data agar lebih mudah dipahami. Dengan teknologi visual penyajian data ilmiah yang kompleks dapat ditampilkan dalam bentuk yang lebih mudah diterima, penyajian data secara visual memiliki karakter sebagai berikut :

a. *Interactivity*, yaitu data yang disajikan dapat dengan mudah untuk diolah.
b. *Multidimensional*, yaitu menyajikan informasi dari berbagai dimensi yang di bentuk kedalam satu dimensi.
c. *Visibility*, yaitu penyajian data secara jelas baik itu dalam bentuk grafik, gambar, kurva, animasi dan pola yang dapat divisualkan.

B. Penelitian Terdahulu

Penelitian (Nguyen & Armitage, 2008) klasifikasi trafik berdasarkan port number dari aplikasi, tetapi ada beberapa aplikasi yang menggunakan port yang sama sehingga sulit untuk mengelompokkan trafik. Penelitian yang di lakukan oleh (Jesudasan, Branch, & But, 2010) , (Alshammari & Zincir-Heywood, 2010) menggunakan *machine learning* (ML) menggunakan *artificial intelligent* untuk mengklasifikasi IP trafik dengan cara mengextrak infromasi dari fitur aplikasi, hal yang sama juga dilakukan oleh (Soysal & Schmidt, 2010) menjelaskan algoritma ML mempunyai kecepatan yang baik untuk klasifikasi trafik, namun (Gu, Wang, & Ji, 2010) menyebutkan klasifikasi dengan menggunakan metode ML hanya untuk jaringan yang *offline* dan tidak mampu untuk klasifikasi trafik untuk jaringan yang online.

Salah satu metode yang banyak dipakai dalam mengklasifikasi trafik adalah *Deep Packet Inspection* (DPI). (Bujlow, Carela-Español, & Barlet-Ros, 2015) membandingkan beberapa tool DPI dalam mengklasifikasi trafik, penelitian yang lain menggunakan DPI adalah (Oklilas & Tasmi, 2017) mampu mengklasifikasi trafik pada jaringan wifi.

(Fan & Liu, 2017) Klasifikasi trafik jaringan dengan metode *Support Vector Machine (SVM)* dan K-Mean dan membandingkan nilai akurasi dari kedua metode tersebut. Hasil dari penelitian ini klasifikasi dengan metode *Support Vector Machine (SVM)* lebih akurat dari pada metode *K-Mean*, dan *p*engaturan model dan pemilihan fitur yang tepat dapat meningkatkan akurasi. Penelitan yang dilakukan oleh (Wang, Zhang, Ye, & Du, 2015) menggunakan *SVM* dan *Artificial Bee Colony*

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

(ABC) Algorithm identifikasi trafik *Peer to Peer*. Selajutnya penelitian (Wu, Dong, Yang, & Tang, 2018) Klasifikasi trafik video berdasarkan penggunaan memori, latensi dan akurasi. dengan *SVM, Bayes* dan *K-NN*.

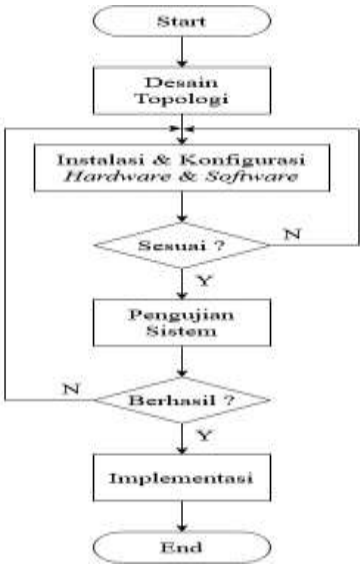
C. Metodologi Penelitian

Terdapat tahapan penelitian secara umum yang merupakan panduan dari pelaksanaan penelitian untuk mencapai tujuan penelitian. Tahapan penelitian ini adalah sebagai berikut : (i). Tahap Perancangan sistem merupakan perancangan topologi, installasi perangkat keras dan perangkat lunak yang akan digunakan dalam penelitian ini, (ii). Tahap Pengumpulan data merupakan proses *Sniffing* pada jaringan yang menjadi objek penelitian, (iii). Tahap *Feature Extraction* merupakan tahapan untuk mendapatkan informasi dari hasil pengumpulan data, (iv). Tahap proses klasifikasi merupakan penerapan metode *Support Vector Machine (SVM)* untuk klasifikasi trafik jaringan, (v). Tahap analisis hasil, merupakan analisa dari hasil klasifikasi.

1. Perancangan Sistem

Tahapan perancangan sistem yang merupakan tahapan mempersiapkan rancangan topologi, instalasi perangkat keras dan perangkat lunak yang akan dipergunakan dalam penelitian, adapun langkah-langkah yang akan dilakukan pada tahap ini adalah sebagai berikut:

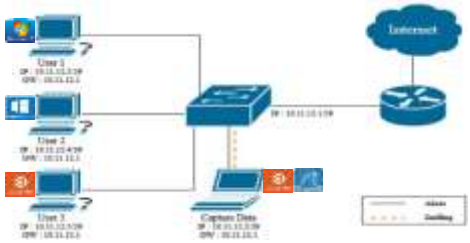
a. Menentukan topologi yang sesuai dengan penelitian yang akan dilakukan.
b. Instalasi dan konfigurasi perangkat keras dan perangkat lunak seperti, (i). Kofigurasi *IP Address*, (ii). Konfigurasi *Network Address Translation (NAT)*, (iii). Konfigurasi *Routing*, (iv). Konfigurasi *DNS Server* dan (v). Konfigurasi *DHCP Server*.
c. Melakukan pengetesan seluruh perangkat dan memastikan seluruh sistem berjalan dengan baik sesuai yang direncanakan, untuk pengetesan yang dapat dilakukan adalah melakukan *Ping* ke sistem baik dari dalam maupun dari luar, tes *Browsing* ke internet dan tes akses ke sistem melalui *Remote* dengan *SSH*.
d. Memastikan seluruh sistem telah berjalan sesuai dengan baik, apabila terdapat kesalahan atau sistem tidak berjalan dengan baik maka silahkan melakukan pengecekan pada langkah 2.



Gambar 6. Tahapan dalam Perancangan Sistem

2. Perancangan Topologi

Penelitian ini dilakukan di *Network Operation Center (NOC)* Universitas Indo Global Mandiri dan yang menjadi objek penelitian adalah jaringan LAN Universitas Indo Global Mandiri.



Gambar 7. Topologi Pengumpulan Data (BPT UIGM. 2020)

Gambar 5 merupakan topologi yang diterapkan dalam pengumpulan data adalah topologi *Star*, sebagai pusat distribusi dari topologi tersebut adalah sebuah perangkat *L2 Switch* yang dapat di *Management* yang berfungsi sebagai *Port Mirroring* untuk mengumpulkan trafik data yang melintas. Pada pengumpulan data ini terdapat 3 buah *PC* sebagai user yang akan melakukan *Browsing* ke internet dan 1 buah laptop yang berfungsi untuk mengumpulkan data dengan teknik *Sniffing*.

3. Pengumpulan Data

Pada tahap ini akan dilakukan pengumpulan data untuk pembuatan dataset yang akan digunakan dalam penelitian, pengambilan dataset dilakukan dengan teknik *Sniffing* dengan menggunakan aplikasi *Wireshark*. Skenario pengambilan dataset seluruh *User* akan melakukan akses ke internet melalui *Web Browser*, seluruh trafik yang melalui *L2 Switch* akan diduplikasi pada laptop yang digunakan untuk *Sniffing*. Salah satu *User* akan mengakses kebeberapa aplikasi yang telah ditentukan antara lain : *Youtube* dan *SoundCloud*.

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

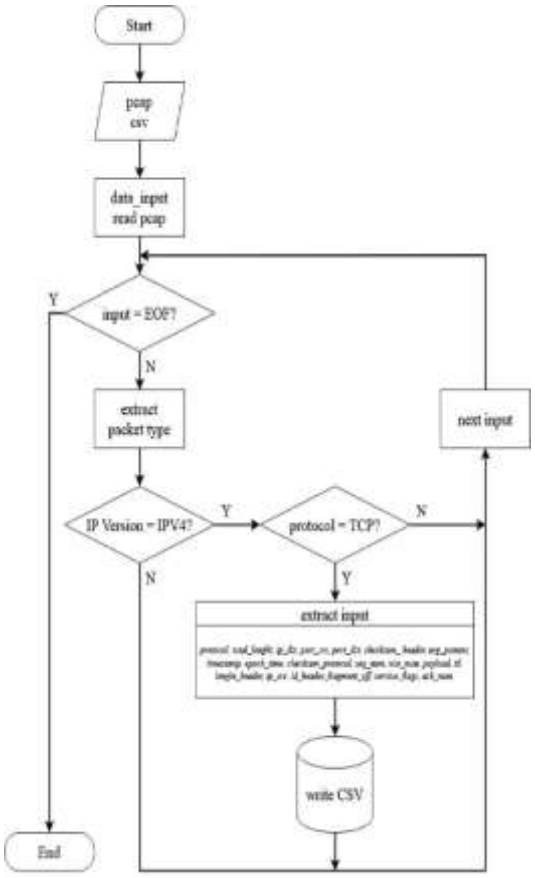
Sedangkan *User* lainnya melakukan akses secara bebas ke internet.

Tabel Error! No text of specified style in document.**.**
Tahapan Pengumpulan Data

Tahapan	Keterangan
Persiapan	Pengambilan dataset dilakukan selama 12 Menit. Proses *Sniffing* dilakukan secara *Real Time* di Laptop yang berfungsi mengumpulkan data.
Skenario 1	Durasi 2 menit, *User* 1 melakukan akses ke Youtube, *User* 2 dan 3 melakukan akses secara bebas ke internet selain akses ke Youtube.
Skenario 2	Durasi 2 menit, *User* 1 melakukan akses ke Sound Cloud, *User* 2 dan 3 melakukan akses secara bebas ke internet selain akses ke Sound Cloud.
Skenario 3	Durasi 2 menit, *User* 2 melakukan akses ke Youtube, *User* 1 dan 3 melakukan akses secara bebas ke internet selain akses ke Youtube.
Skenario 4	Durasi 2 menit, *User* 2 melakukan akses ke Sound Cloud, *User* 1 dan 3 melakukan akses secara bebas ke internet selain akses ke Sound Cloud.
Skenario 5	Durasi 2 menit, *User* 3 melakukan akses ke Youtube, *User* 1 dan 2 melakukan akses secara bebas ke internet selain akses ke Youtube.
Skenario 6	Durasi 2 menit, *User* 3 melakukan akses ke Sound Cloud, *User* 1 dan 2 melakukan akses secara bebas ke internet selain akses ke Sound Cloud.

4. *Feature Extraction*

Pada tahap *Feature Extraction* ini dataset yang telah dibuat dengan format file *Packet Capture (pcap)* akan diubah menggunakan program dengan Bahasa pemrograman *Python* untuk mendapatkan informasi atribut-atribut sebuah paket data yang akan disimpan menjadi file dengan format *Comma Separated Value (CSV).* Gambar 6 menjelaskan *Flowchart* program untuk proses *Feature Extraction*, dari *Flowchart* tersebut dapat ditulis dalam bentuk *Pseudocode*



Gambar 8. *Flowchart Program Feature Extraction*

```
input : pcap
  output : csv
  Begin
    read pcap
    data_input ← read pcap
    while data_input != EOF do
      extract packet type
      if IP Version = IPV4 then
        if protocol = TCP then
          write extracted data to CSV
        else
          write other to CSV
        end if
      end if
    end while
  end
```

Gambar 9. *Pseudocode program Feature Extraction*

Gambar 7 *Pseudocode* program *Feature Extraction*, pada program *Feature Extraction* terdapat 2 parameter, pertama parameter F untuk menandakan *File* masukan yang berupa *File PCAP* dan parameter kedua O untuk memberikan nama *File* keluaran dengan format *CSV.* Atribut yang didapat dari hasil *Feature Extraction* Dari *Feature Extraction* menghasilkan sebanyak 21 atribut yang nantinya akan dipilih atribut terbaik dan digunakan dalam *Modeling Data.*

Data *Raw* yang telah dilakukan proses *Feature Extraction* selanjutnya dilakukan pembersihan data, pemilihan fitur yang akan dijadikan atribut dilakukan

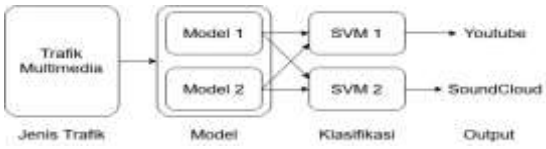
ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

secara manual. Tidak semua fitur akan dipakai, hanya fitur-fitur yang memang menjadi ciri dari pola trafik yang akan dipakai dan dijadikan data set

5. Proses Klasifikasi

Pada penelitian ini trafik yang akan diklasifikasikan adalah trafik dengan kategori *Multimedia* seperti Youtube dan Sound Cloud, metode klasifikasi yang digunakan adalah *Support Vector Machine (SVM)*. Data trafik yang akan diklasifikasikan akan diambil secara *Real Time* yang bertempat di *Network Operation Center (NOC)* Universitas Indo Global Mandiri, waktu pengambilan data pada saat kondisi puncak penggunaan jaringan yang didapat dari data penggunaan jaringan dengan aplikasi *MikroTik Graphing*

Untuk mengidentifikasi trafik multimedia pada penelitian ini akan digunakan dua model yaitu model 1 sebagai pendeteksi trafik Youtube dan model 2 sebagai pendeteksi trafik SoundCloud. Jenis trafik akan dikenali polanya melalui proses klasifikasi *SVM* menggunakan metode *One Vs Other.*



Gambar 10. *Arsitektur SVM dengan Metode One vs Other*

Gambar diatas menjelaskan arsitektur *SVM* dengan metode *One vs Other,* langkah-langkah dalam meng-klasifikasikan trafik multimedia dapat diuraikan sebagai berikut :

1. Tahap pertama akan ditentukan jumlah kelas *SVM Biner,* kemudian memetakan data dari *Input Space* ke dalam *Feature Space* menggunakan kernel *Radial Basis Function (RBF)* seperti pada persamaan dibawah ini

$$K(\vec{x}, \vec{y}) = exp(-\gamma \| \vec{x} - \vec{y} \|^2)$$

Kemudian tentukan nilai *Support Vector* dari nilai $\alpha \neq 0$ dengan menghitung nilai $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$ dari *Quadrating Program,* maka akan diperoleh data X_1 yang berkorelasi dengan $\alpha \neq 0$ sebagai *Support Vector.*

2. Tahapan pengujian, yaitu menentukan jumlah kelas *SVM* biner, kemudian memetakan data dari *Input Space* ke dalam *Feature Space* menggunakan kernel *RBF*, dan menghitung fungsi pengambilan keputusan terakhir.

```
Deklarasi:
        tr1, tr2, trDetected: String
        isPredicted: Boolean
        i, predicted: Integer

    Deskripsi:
        for (integer i → 0; i < 2; i++) do
            File Copy (TRAIN_FILE[i], TEST_FILE, true);
            Write = read (TEST_FILE, true);
            Write ();
            Write ("0 1:" + tr1 + "2:" + tr2);
        Close();
            isPredicted[i] → predicted == 1 ? true : false;
        end for
            if (isPredicted[0] and isPredicted[1]) then
                    trDetected → "UNPREDICTED";
                else if (isPredicted[0] and not
isPredicted[1]) then
                        trDetected → "YOUTUBE";
                else if (not isPredicted[0] and
isPredicted[1]) then
                        trDetected → "SOUNDCLOUD";
                    else
                            trDetected →
"UNPREDICTED";
                    end if
                end if
                end if
            end if
```

Gambar 11. *Pseudocode Program SVM*

Pseudocode Program *SVM*, data akan diproses dengan algoritma *SVM* untuk memprediksi jenis trafik.

2. Hasil Dan Pembahasan

A. Dataset

Pada penelitian ini kami beberapa tahap seperti yang telah dijelasakan pada point tiga. Penelitian kami hanya mengambil data *streaming* yaitu: *Youtube* dan *SoundCloud.* Tahap awal penelitian ini adalah pengumpulan data yang dilakukan dengan teknik *sniffing* menggunakan aplikasi *wireshark*, pengumpulan data dilakukan dengan enam percobaan yang masing-masing percobaan akan dibatasi lebih kurang dua menit.

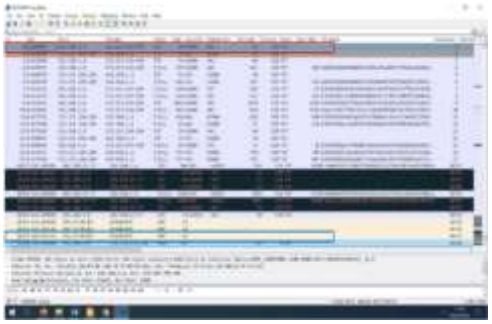
Tabel Error! No text of specified style in document.**.** Hasil Pengumpulan Data berdasarkan Protokol Komunikasi

Perc-ke	Jumlah Paket	File-Size (MB)	*TCP* (%)	*UDP* (%)	Drop (%)	Trafik
1	38767	41,190	99,7	0,1	0,2	Youtube
2	5941	6,511	97,6	1,1	1,3	SoundCloud
3	35120	39,261	99,7	0,2	0,1	Youtube
4	9524	9,557	97,5	1,2	1,3	SoundCloud
5	20882	21,861	99,5	0,2	0,3	Youtube
6	5904	6,730	97,7	1,1	1,2	SoundCloud

Tabel 4 menampilkan hasil dari percobaan berdasarkan protokol komunikasi, dari enam percobaan pengumpulan data. Berdasarkan data yang dihasilkan terdapat perbedaan untuk jenis trafik *Youtube* dan *SoundCloud* yaitu: (i) Dengan waktu yang sama jumlah

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

paket data trafik *Youtube* lebih banyak dibandingkan jumlah paket data trafik *SoundCloud*, (ii) Berdasarkan protokol komunikasi trafik *Youtube* dan *SoundCloud* komunikasi yang terjadi paling banyak menggunakan protocol *TCP* untuk mengakses dua aplikasi tersebut, dan (iii) tidak terjadi perbedaan hasil yang didapatkan secara keseluruhan antara akses yang menggunakan *Webbrowser Mozilla Firefox* dan *Google Chrome*



Gambar 12. *Capture trafik dengan wireshark*

Berdasarkan hasil pengamatan hasil *capture* paket data yang telah dilakukan telah kami sesuai dengan perancangan topologi jaringan, dimana dapat dilihat *IP Address* yang didapat dari hasil *Capture* sesuai dengan *IP Address* yang terdapat pada *Client.* Selain dari topologi jaringan data dapat juga dilihat dari waktu awal dan waktu berakhirnya proses *Capture* pada setiap percobaan yang telah dilakukan dan jumlah paket yang berhasil di *Capture.*

B. Feature Extraction

Untuk mendapatkan dataset secara *real time* perlu dilakukan pengumpulan data secara *real time* dan dilakukan proses ekstraksi fitur untuk mendapatkan informasi dari paket data yang melalui jaringan.

Gambar 11 menampilkan hasil ekstraksi paket data, terdapat beberapa fitur yang dihasilkan antara lain : pada protokol *UDP* dan *TCP* sedangkan pada layer aplikasi ada beberapa *service* seperti *telnet, http, ftp,* dan *snmp*

Gambar 13. Hasil *Feature Extraction*

Sebagai dasar untuk validasi data hasil pengujian kami, maka disini membandingkan hasil *raw data (pcap)* dan hasil *feature extraction* seperti ditampilkan pada gambar 12

Gambar 14. *Validasi Hasil Feature Extraction*

Gambar diatas menampilkan validasi data hasil *Feature Extraction* dengan *Raw Data* berdasarkan protokol komunikasinya

C. Proses Klasifikasi dan Visualisasi

Sebelum melakukan proses klasfikasi terlebih dahulu kami melakukan pembacaan di memberi label pada setiap jenis trafik yang di capture. Gambar 13 menjelaskan hasil ekstraksi dari dataset yang didalam *Payload*nya terdapat trafik yang mengandung aliran data dari Soundcloud. Berdasarkan *Payload* tersebut fitur-fitur pada *Payload* tersebutlah yang nantinya dipakai sebagai ciri trafik Soundcloud pada proses klasifikasinya

Gambar 15. *Payload* yang mengandung trafik *Soundcloud.*

Pada bagian selanjutnya kami mencoba melakukan klasfikasi data dari raw data yang di hasilkan dari capture, pada kegiatan ini klasfikasi trafik berdasarkan *port number* dan jumlah paket data. Hasil klasifikasi yang didapatkan dalam bentuk persentase dari beberapa kategori aplikasi adalah sebagai berikut

Tabel 5. Hasil Klasifikasi

Kategori	Persentase
Games	0,4%
Mail	0,2%
Messenger	0,3%
Multimedia	0,4%

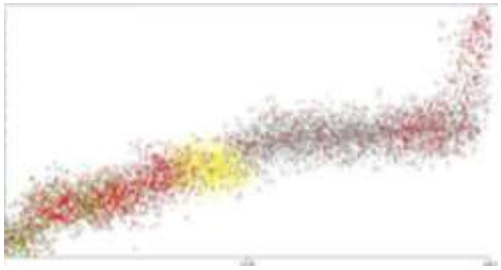
ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

P2P	0,3%
Services	0,1%
Web	82,8%
Other	15,5%

Tabel 5 menampilkan hasil klasifikasi yang terdiri dari beberapa kategori dan besar persentasenya, *port number* dan jumlah paket digunakan sebagai fitur utama untuk mengenali setiap jenis aplikasi yang lewat didalam jaringan. Dari hasil klasifikasi akses terbesar merupakan trafik *Web*, oleh karena itu perlu dilakukan klasifikasi secara mendalam untuk trafik *Web* berdasarkan port. Hasil dari klasifikasi tersebut dapat dilihat pada gambar 14 yang menjelaskan hasil klasifikasi berdasarkan *Service* dan protokol yang digunakan pada aplikasi berbasis *Web* yang terbanyak yaitu *HTTPS.*

Gambar 16. Hasil Klasifikasi Trafik

Berdasarkan hasil pengamatan pada dataset yang telah diekstraksi terdapat beberapa fitur dengan nilai dominan terhadap service yang dipergunakan dan ditampilkan dalam bentuk visualisasi. Pada penelitian ini kami mencoba penampilkan hasil visualisasi dari raw data yang didapat data *Real Time* yang sebelumnya telah di *Capture* pada jaringan eksisting UIGM, *Raw Data* dari hasil pengumpulan data selanjutnya diolah untuk menampilkan data tersebut dalam bentuk visual yang dapat dilihat pada gambar 14 dan 15



Gambar 17. *Visualisasi Raw Data*

4. Kesimpulan

Untuk mendapatkan hasil klasifikasi trafik yang akurat diperlukan proses *Feature Extraction,* proses ekstraksi fitur secara mendalam akan menghasilkan banyak informasi dari paket data yang dapat digunakan sebagai atribut dalam proses klasifikasi. trafik *Web* merupakan trafik tertinggi sehinggan harus diatur prioritasnya menjadi yang paling utama dalam jaringan UIGM. Pada penelitian selanjutnya proses *feature extraction* sebaiknya menggunakan metode-metode yang dapat mengukur tingkat relasi antar fitur sehingga didapatkan atribut yang kuat. Penelitian selanjutnya dapat lebih menekankan pada model dan rule klasifikasi yang dibangun sehingga akurasi yang didapat akan lebih baik

Daftar Pustaka

Alshammari, R., & Zincir-Heywood, A. N. (2010). An investigation on the identification of voip traffic: Case study on Gtalk and Skype. *Proceedings of the 2010 International Conference on Network and Service Management, CNSM 2010*, 310–313. https://doi.org/10.1109/CNSM.2010.5691210

Awad, H., Ibrahim, H., Radhi, O., Al, A., Al-namari, M. A., & Mohamedali, G. (2016). Internet Traffic Classification using Machine Learning Approach : Datasets Validation Issues. *Proceedings of 2016 Conference of Basic Sciences and Engineering Studies, SGCAC 2016*, 158–166. https://doi.org/10.1109/SGCAC.2016.7458022

Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., … Caicedo, O. M. (2018). A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications*, Vol. 9, pp. 1–99. https://doi.org/10.1186/S13174-018-0087-2

Bujlow, T., Carela-Español, V., & Barlet-Ros, P. (2015). Independent comparison of popular DPI tools for traffic classification. *Computer Networks*. https://doi.org/10.1016/j.comnet.2014.11.001

Byun, H., & Lee, S. W. (2003). A survey on pattern recognition applications of support vector machines. *International Journal of Pattern Recognition and Artificial Intelligence*, *17*(3), 459–486. https://doi.org/10.1142/S0218001403002460

Fan, Z., & Liu, R. (2017). Investigation of machine learning based network traffic classification. *Proceedings of the International Symposium on Wireless Communication Systems*. https://doi.org/10.1109/ISWCS.2017.8108090

Gu, R., Wang, H., & Ji, Y. (2010). Early Traffic Identification Using Bayesian Networks. *Communication*, 564–568.

Hao, S., Hu, J., Liu, S., Song, T., Guo, J., & Liu, S. (2015). Improved SVM method for internet traffic classification based on feature weight learning. *ICCAIS 2015 - 4th International Conference on Control, Automation and Information Sciences*, 102–106. https://doi.org/10.1109/ICCAIS.2015.7338641

Jesudasan, R. N., Branch, P., & But, J. (2010). *Generic Attributes for Skype Identification Using Machine Learning*. (August), 1–7.

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

Kong, L., Huang, G., Wu, K., Tang, Q., & Ye, S. (2018). Comparison of Internet Traffic Identification on Machine Learning Methods. *2018 International Conference on Big Data and Artificial Intelligence (BDAI)*, 38–41. https://doi.org/10.1109/BDAI.2018.8546682

Liu, C. C., Chang, Y., Tseng, C. W., Yang, Y. T., Lai, M. S., & Chou, L. Der. (2018). SVM-based Classification Mechanism and Its Application in SDN Networks. *2018 10th International Conference on Communication Software and Networks, ICCSN 2018*, 45–49. https://doi.org/10.1109/ICCSN.2018.8488219

Networks, C., Dhote, Y., Agrawal, S., & Deen, A. J. (2015). A Survey on Feature Selection Techniques for Internet Traffic Classification. *Proceedings - 2015 International Conference on Computational Intelligence and Communication Networks, CICN 2015*, 1375–1380. https://doi.org/10.1109/CICN.2015.267

Nguyen, T. T. T., & Armitage, G. (2008). A Survey of Techniques for Internet Traffic Classification Using Machine Learning. *Communications Surveys & Tutorials, IEEE*, *10*(4), 56–76. https://doi.org/10.1109/SURV.2008.080406

Oklilas, A. F., & Tasmi. (2017). Monitoring and Indentification Packet in Wireless With Deep Packet Inspection Method. *International Conference on Recent Trends in Physics 2016 (ICRTP2016) IAES International Conference on Electrical Engineering, Computer Science and Informatics*, *365*, 011001. https://doi.org/10.1088/1742-6596/365/1/011001

Parziale, L., Britt, D. T., Davis, C., Forrester, J., Liu, W., Matthews, C., & Rosselot, N. (2006). TCP/IP Tutorial and Technical Overview. In *TCP/IP Tutorial and Technical Overview* (8th ed., Vol. 8). IBM Corporation.

Robertazzi, G. T. (2017). *Introduction To Computer Networking*. https://doi.org/10.1007/978-3-319-53103-8

Shafiq, M., Yu, X., Laghari, A. A., Yao, L., Karn, N. K., & Abdessamia, F. (2017). Network Traffic Classification techniques and comparative analysis using Machine Learning algorithms. *2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings*, 2451–2455. https://doi.org/10.1109/CompComm.2016.7925139

Siqueira, I. G., Ruiz, L. B., & Loureiro, a. a. F. (2007). Coverage area management for wireless sensor networks. *Internation Journal of Network Management*, (October 2005), 17–31. https://doi.org/10.1002/nem

Soysal, M., & Schmidt, E. G. (2010). Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison. *Performance Evaluation*, *67*(6), 451–467. https://doi.org/10.1016/j.peva.2010.01.001

Wang, C., Zhang, H., Ye, Z., & Du, Y. (2015). A peer to peer traffic identification method based on support vector machine and artificial bee colony algorithm. *Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2015*, *2*(September), 982–986. https://doi.org/10.1109/IDAACS.2015.7341451

Wu, Z., Dong, Y., Yang, L., & Tang, P. (2018). A New Structure for Internet Video Traffic Classification Using Machine Learning. *Proceedings - 2018 6th International Conference on Advanced Cloud and Big Data, CBD 2018*, (61271233), 322–327. https://doi.org/10.1109/CBD.2018.00064

Yang, J., Narantuya, J., & Lim, H. (2019). Bayesian Neural Network Based Encrypted Traffic Classification using Initial Handshake Packets. *Proceedings - 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume, DSN-S 2019*, 19–20. https://doi.org/10.1109/DSN-S.2019.00015

Zamfir, S., Balan, T., Sandu, F., & Costache, C. (2016). Solutions for deep packet inspection in industrial communications. *IEEE International Conference on Communications*, *2016-Augus*, 153–158. https://doi.org/10.1109/ICComm.2016.7528337