

Klasifikasi Data Trafik Internet Menggunakan Metode *Bayes Network* (Studi Kasus Jaringan Internet Universitas Semarang)

Rastri Prathivi

Jurusan Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang

Jl. Soekarno-Hatta Semarang 50196

Telp. (024) 6702757

E-mail: vivi@usm.ac.id

Abstract— Pemakaian internet merupakan kebutuhan yang penting yang mendukung kinerja dan aktivitas di kampus. Bagian yang terpenting dari infrastruktur internet yang difasilitasi oleh kampus adalah tersedianya *bandwidth* yang cukup untuk kelancaran trafik data melalui internet. Metode klasifikasi menggunakan *Bayes Network* ini memanfaatkan metode klasifikasi yang dimiliki oleh data *mining* untuk diterapkan pada data trafik jaringan internet.

Penelitian ini bertujuan untuk mengklasifikasi data pemakaian internet sehingga dari klasifikasi tersebut dapat diketahui *destination network*, *protocol* dan lebar *bandwidth* yang banyak diakses pada waktu tertentu.

Data trafik internet diambil melalui *software Wireshark*. Sedangkan pengolahan data dan proses pengklasifikasian data trafik internet diolah dengan Weka.

Keywords—*Bayes Network*, *Klasifikasi*, *Trafik*

I. PENDAHULUAN

Pemakaian internet merupakan kebutuhan yang penting yang mendukung kinerja dan aktivitas di kampus. Bagian yang terpenting dari infrastruktur internet yang difasilitasi oleh kampus adalah tersedianya *bandwidth* yang cukup untuk kelancaran trafik data melalui internet. Seringkali besar *bandwidth* yang tersedia dirasakan kurang, terutama pada jam-jam sibuk dan perkuliahan yang aktif.

Metode klasifikasi data terhadap trafik internet bisa digunakan untuk mengetahui data pemakaian internet yang ada sehingga pengaturan *bandwidth* dapat dilakukan agar tercapai suatu koneksi internet yang handal dan stabil. Secara umum metode klasifikasi trafik dapat digolongkan ke dalam metode *Port-Based*, *Payload-Based*, *Protocol Behavior or Heuristic Based Classification* dan Analisa Klasifikasi berdasarkan data statistika.

Metode klasifikasi menggunakan *Bayes Network* ini memanfaatkan metode klasifikasi yang dimiliki oleh data *mining* untuk diterapkan pada data trafik jaringan internet.

Penelitian yang berkaitan dengan klasifikasi data trafik internet atau jaringan pernah dilakukan oleh beberapa penelitian antara lain:

Nguyen dan Armitage mengusulkan klasifikasi trafik alamat IP dengan metode Naive Bayes. Kriteria utama yang membedakan antara teknik klasifikasi adalah akurasi prediksi (yaitu, seberapa akurat teknik atau model yang digunakan dapat membuat keputusan saat menyajikan data sebelumnya yang tak terlihat). Sejumlah metrik dapat digunakan untuk mengekspresikan akurasi klasifikasi.

Gowsalya dan Miruna mengusulkan metode klasifikasi trafik jaringan dengan *Naive Bayes* menggunakan korelasi arus informasi. *Naive Bayes* estimator diterapkan dalam aplikasi untuk mengkategorikan lalu lintas dalam jaringan. Dalam metode yang diusulkan oleh Gowsalya dan Miruna, arus lalu lintas dijelaskan menggunakan diskritisasi fitur statistik dan korelasi arus informasi yang dimodelkan dengan bag-of-aliran (BOF). Pendekatan parametrik baru untuk klasifikasi lalu lintas, yang dapat meningkatkan klasifikasi kinerja secara efektif dengan memasukkan informasi berkorelasi ke dalam proses klasifikasi.

Mengacu pada penelitian sebelumnya maka dalam penelitian ini penulis menggunakan metode *Bayes Network*. Penelitian ini bertujuan untuk mengklasifikasi data pemakaian internet sehingga dari klasifikasi tersebut dapat diketahui *destination network*, *protocol* dan lebar *bandwidth* yang banyak diakses pada waktu tertentu.

II. METODE PENELITIAN

A. Identifikasi Variabel Penelitian

Variabel penelitian di dalam penelitian ini adalah:

1. *Destination* menunjukkan data tujuan dari pemakaian trafik internet.
2. *Protocol* menunjukkan data protokol yang sering diakses.
3. *Length* menunjukkan data lebar *bandwidth* yang digunakan.

B. Metode Pengumpulan Data

1. Data Primer
Data yang digunakan sebagai data primer dalam penelitian ini adalah data pemakaian trafik internet

yang diambil dari data jaringan internet Universitas Semarang.

2. Data Sekunder

Sedangkan data sekunder dalam penelitian ini berasal dari studi pustaka yang berkaitan dengan judul penelitian.

C. Metode Klasifikasi Data

1. Preprocessing

Pada proses data *preprocessing* ini data pemakaian trafik internet di Universitas Semarang diambil menggunakan *software Wireshark*. Data yang diambil kemudian disimpan dalam format file *.csv*.

Setelah diperoleh data pemakaian trafik internet, selanjutnya adalah proses *preprocessing* dengan *software Weka*. Caranya adalah:

- Masuk ke *applications explorer*.
- Pada menu *preprocess*, pilih tab *Open file*.
- Pilih file *.csv*.

2. Klasifikasi Data dengan Bayes Network

Setelah data melalui proses *preprocessing* maka dilakukan proses klasifikasi. Untuk melakukan proses klasifikasi caranya adalah:

- Pilih tab *Classify*.
- Pada bagian tab *options*, pilih *Use training set*.
- Klik *Start* untuk melakukan proses klasifikasi.

III. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini dapat dijelaskan pada gambar-gambar berikut ini:

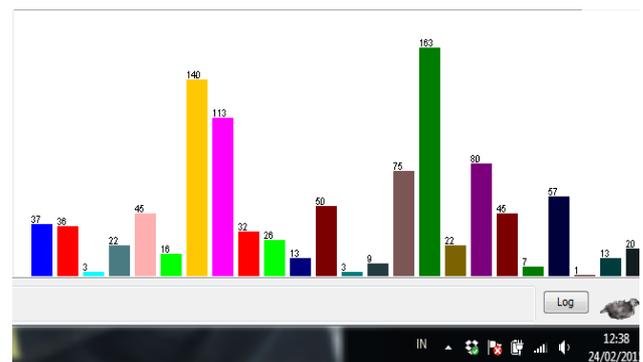
A. Hasil Proses Preprocessing

Proses *Preprocessing* menghasilkan data jumlah total dari *destination network*, *protocol* dan lebar *bandwidth* yang banyak diakses dalam trafik jaringan.

Tabel 1. Jumlah Total IP *Destination*

NO.	IP DESTINATION	COUNT
1	192.168.223.1	37
2	192.168.223.100	36
3	HonHaiPr_76:8a:81	3
4	HonHaiPr_0f:e9:bf	22
5	255.255.255.255	45
6	192.168.223.255	16
7	ff02::1:3	140
8	224.0.0.252	113
9	ff02::16	32
10	224.0.0.22	26
11	ff02::1:2	13
12	Broadcast	50

NO.	IP DESTINATION	COUNT
13	ff02::1:ffe2:713b	3
14	ff02::2	9
15	239.255.255.250	75
16	192.168.150.255	163
17	172.16.1.1	22
18	192.168.150.101	80
19	ff02::c	45
20	192.168.150.1	7
21	169.254.255.255	57
22	Cisco-Li_49:c8:c7	1
23	202.152.130.42	13
24	74.125.68.102	20



Gambar 1. Grafik Jumlah Total *Destination Network*

Tabel 1 dan gambar 1 memperlihatkan jumlah total dari *Destination Network* yang banyak diakses dalam jaringan internet. *Destination Network* berupa alamat- alamat web dan alamat jaringan yang banyak diakses pada hari Jumat tanggal 13 Februari 2015.

Tabel 2. Jumlah Total *Protocol*

NO.	PROTOCOL	COUNT
1	DNS	80
2	ICMP	35
3	ARP	76
4	DB-LSP-DISC	24
5	DHCP	29
6	LLMNR	253
7	ICMPv6	44
8	IGMPv3	26
9	DHCPv6	13
10	NBNS	207
11	SSDP	130
12	BROWSER	23


```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	DNS
	1	0.001	0.972	1	0.986	1	ICMP
	1	0	1	1	1	1	ARP
	1	0	1	1	1	1	DB-LSP-DISC
	0.966	0	1	0.966	0.982	1	DHCP
	1	0	1	1	1	1	LLMNR
	1	0	1	1	1	1	ICMPv6
	1	0	1	1	1	1	IGMPv3
	1	0	1	1	1	1	DHCPv6
	1	0	1	1	1	1	NBNS
	1	0	1	1	1	1	SSDP
	1	0	1	1	1	1	BROWSER
	0.976	0.001	0.976	0.976	0.976	1	TCP
	0.5	0	1	0.5	0.667	1	HTTP
	1	0	1	1	1	1	UDP
	1	0.001	0.976	1	0.988	1	TLSv1.2
Weighted Avg.	0.997	0	0.997	0.997	0.997	1	

Gambar 7. Hasil Klasifikasi Data *Protocol*
 Gambar 7 memperlihatkan akurasi per *class* dari klasifikasi pada *protocol*.

```

=== Confusion Matrix ===

```

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	<-- classified as	
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a = DNS
0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = ICMP
0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c = ARP
0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	d = DB-LSP-DISC
0	1	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	e = DHCP
0	0	0	0	0	253	0	0	0	0	0	0	0	0	0	0	0	f = LLMNR
0	0	0	0	0	0	44	0	0	0	0	0	0	0	0	0	0	g = ICMPv6
0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	h = IGMPv3
0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	i = DHCPv6
0	0	0	0	0	0	0	0	0	207	0	0	0	0	0	0	0	j = NBNS
0	0	0	0	0	0	0	0	0	0	130	0	0	0	0	0	0	k = SSDP
0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	l = BROWSER
0	0	0	0	0	0	0	0	0	0	0	0	40	0	0	1	0	m = TCP
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	n = HTTP
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	o = UDP
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40	p = TLSv1.2

Gambar 8. Matriks Klasifikasi *Protocol*
 Dari gambar 8 terlihat bahwa *protocol* yang paling banyak diakses adalah LLMNR sebanyak 253 data

IV. KESIMPULAN DAN SARAN

Dari penelitian ini dapat diambil beberapa kesimpulan yaitu:

1. Metode *Bayes Network* menghasilkan klasifikasi trafik jaringan sesuai dengan *destination network* dan *protocol*.
2. Dari klasifikasi tersebut terlihat pemakaian dari alamat jaringan tau internet dan *protocol* jaringan yang paling banyak diakses.

Saran untuk penelitian ini yaitu:

1. Dapat dikembangkan metode klasifikasi trafik jaringan yang lain untuk mendeteksi gangguan jaringan.
2. Metode klasifikasi jaringan dapat dioptimalkan dengan menggabungkan metode lain untuk optimasi data.

REFERENSI

- [1] R.S.Anu Gowsalya, S. Miruna Joe Amali, "Naive Bayes Based Network Traffic Classification Using Corellation Information," IJARCSSE Volume 4. Issue 3ty,pp 8-14, March 2014. (*references*).
- [2] T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," Communication Surveys Tutorials., vol. 10, no. 4, pp. 56–76, 4th Quarter 2008.
- [3] Michal Horny, "Bayesian Network", Technical Report No. 5 Boston University, 18 April 2014.
- [4] Adi Suryaputra, Febrillian Samopa dan Bekti Cahyo Hindayanto, "Klasterisasi dan Analisis Trafik Internet Menggunakan Fuzzy C Mean dengan Ekstraksi Fitur Data", Jurnal Informatika, Vol. 12, No. 1, pp 33-39, Mei 2014.
- [5] http://en.wikipedia.org/wiki/Bayesian_network. Diakses tanggal 26 Februari 2015.
- [6] http://en.wikipedia.org/wiki/Traffic_classification. Diakses tanggal 26 Februari 2015.