

Klasifikasi Penggunaan Protokol Komunikasi Pada Trafik Jaringan Menggunakan Algoritma *K-Nearest Neighbor*

I Komang Kompyang Agus Subrata¹, I Made Oka Widyantara², Linawati³

Abstract – Network traffic internet is data communication in a network, which is characterized by information from the packet header data. Proper classification of network traffic on a very important notably in terms of architectural design network design, network and network security Risk Management. The analysis of computer network traffic is one way to know the use of the computer network communication protocol, so it can be the basis for determining the priority of Quality of Service (QoS). Basic practice in providing QoS priority is to analyze the data traffic on the network. In this study the classification of the data capture network traffic on though Algorithm K-Nearest Neighbor (K-NN). Applications used to capture network traffic that wireshark application. The results of observations of the dataset and the network traffic through the calculation process using K-NN algorithm obtained a result that classification K-NN has a very high level of accuracy. This is evidenced by the results of calculations which reached 99.14%, ie by calculating $k = 3$.

Intisari—Trafik jaringan internet merupakan lalu lintas komunikasi data dalam suatu jaringan, yang ditandai dengan informasi dari header paket data. Klasifikasi yang tepat terhadap sebuah trafik jaringan sangat penting dilakukan terutama dalam hal disain perancangan arsitektur jaringan, manajemen jaringan dan keamanan jaringan. Analisa terhadap suatu trafik jaringan komputer merupakan salah satu cara mengetahui penggunaan protokol komunikasi jaringan komputer, sehingga dapat menjadi dasar penentuan prioritas *Quality of Service (QoS)*. Dasar didalam pemberian prioritas *QoS* yaitu dengan menganalisa terhadap data trafik jaringan. Pada penelitian ini melakukan klasifikasi terhadap data *capture* trafik jaringan yang di olah menggunakan *Algoritma K-Nearest Neighbor (K-NN)*. Aplikasi yang digunakan untuk *capture trafik jaringan* yaitu aplikasi *wireshark*. Hasil *observasi* terhadap *dataset* trafik jaringan dan melalui proses perhitungan menggunakan *Algoritma K-NN* didapatkan sebuah hasil bahwa klasifikasi *K-NN* memiliki tingkat keakuratan yang sangat tinggi. Hal ini dibuktikan dengan hasil perhitungan yang mencapai nilai 99,14 % yaitu dengan perhitungan $k = 3$.

Kata Kunci —*Network protocol, K-NN, QoS, network capture*

¹Jurusan Teknik Elektro dan Komputer Universitas Udayana, Kampus Bukit Jimbaran, Badung Bali. 80361, Tel. 0361703315 fax: 0361703315; email: Komangkompyang@yahoo.co.id

^{2,3}Program passca Teknik elektro dan komputer Universitas Udayana, Kampus sudirmasn, Denpasar Bali. 80361, Tel. 0361703315 fax: 0361703315; email: oka.widyantara@unud.ac.id, linawati@unud.ac.id

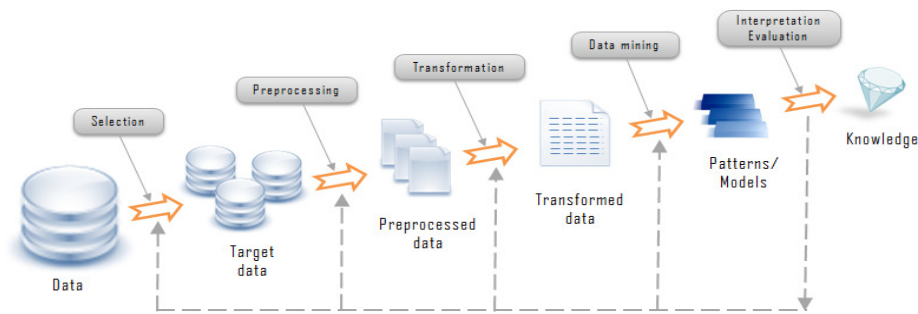
I. PENDAHULUAN

Pada saat ini komunikasi data pada jaringan internet mencapai kemajuan yang sangat pesat, sehingga sudah begitu banyak variasi data yang disebarkan melalui internet, yang dulunya hanya melewati paket-paket data biasa, kini sesuai dengan kebutuhan trafik jaringan internet sudah dilewati paket-paket multimedia seperti audio dan video. Hal ini akan berakibat pada meningkatnya trafik data yang dapat menyebabkan penurunan performansi jaringan terutama pada jaringan yang memiliki *bandwidth* terbatas[1].

Klasifikasi yang tepat terhadap sebuah trafik jaringan internet sangat penting dilakukan terutama dalam hal disain perancangan arsitektur jaringan, manajemen jaringan dan keamanan jaringan. Klasifikasi yang dilakukan yaitu berdasarkan atas banyaknya tipe aktifitas komunikasi, yang diatur oleh protocol-protokol jaringan. Analisa terhadap suatu trafik jaringan adalah salah satu cara mengetahui penggunaan protokol komunikasi jaringan, sehingga dapat menjadi dasar untuk penentuan prioritas suatu trafik jaringan[1].

Terkait dengan klasifikasi trafik internet, beberapa penelitian telah dilakukan dengan menggunakan metode data mining, penerapan algoritma K-NN untuk klasifikasi trafik jaringan telah diajukan oleh[2], Dimana kinerja sistem aplikasi klasifikasi yang dihasilkan cukup baik, karena penerapan algoritma K-NN memiliki kemudahan yaitu tidak ada persyaratan prosedur pelatihan, dan secara alami mampu menangani sejumlah besar kelas. Selanjutnya, penerapan algoritma Naive Bayes oleh[3], juga mampu membuat klasifikasi yang baik dengan mekanisme penentuan Qos berdasarkan protocol-protokol jaringan yang sering digunakan. Selanjutnya Penerapan algoritma K-NN untuk klasifikasi data wine diajukan oleh[4], adalah mengklasifikasi minuman wine menggunakan algoritma *K-NN*, dimana pengklasifikasian *wine* menggunakan dua cara, yaitu pengukuran obyektif dan pengukuran subyektif. Pengukuran secara obyektif dilakukan dengan cara uji laboratorium berdasarkan senyawa yang terkandung dalam *wine* tersebut. Sedangkan untuk pengukuran subyektif dilakukan oleh seorang pakar yang ahli dalam menilai langsung karakteristik *wine*. pengukuran secara subyektif dengan melibatkan pakar yang ahli dalam hal *wine*. Dan yang terakhir dilakukan oleh, adalah buku referensi yang membahas tentang algoritma K-NN. Dalam bukunya menjelaskan bahwa klasifikasi dengan menggunakan algoritma K-NN memiliki tingkat akurasi yang cukup baik. *K-Nearest Neighbor* merupakan metode yang mampu memberikan sebuah keputusan dalam bentuk klasifikasi yang terbagi atas kategori tertentu yang dasar pengklasifikasiannya didapat dari hasil perhitungan dan analisa data latih yang ada.





Gambar 1. Tahapan Data Mining

Pemapanan diata menunjukan bahwa penerapan algoritma *K-Nearest Neighbor* (*K-NN*) adalah solusi yang bisa diterapkan pada mekanisme klasifikasi trafik internet untuk penentuan prioritas layanan QoS. Selanjutnya, paper ini akan mengevaluasi lebih lanjut kinerja penerapan algoritma *K-Nearest Neighbor* untuk klasifikasi trafik jaringan dengan studi kasus pada trafik internet universitas Udayana. Sasarannya adalah meningkatkan kinerja suatu jaringan dan memudahkan *administrator* jaringan didalam manajemen suatu jaringan. Mekanisme klasifikasi yang digunakan adalah mengklasifikasi objek baru berdasarkan atribut dan *training samples* dimana dari data uji yang baru diklasifikasikan berdasarkan jarak terdekat dalam data pelatihan.

Paper ini diorganisasikan sebagai berikut: bab 1 memaparkan tentang latar belakang dan penggambaran umum lingkup penelitian, bab 2 memaparkan tentang tinjauan pustaka, bab 3 memaparkan tentang metodologi penelitian, bab 4 memaparkan tentang hasil dan pembahasan, dan bab 5 memaparkan tentang kesimpulan

II. ANALISIS DATA MINING

A. Data Mining

Han dan Kamber (2006) dalam bukunya yang berjudul “Data Mining Concepts and Techniques” mengatakan, secara singkat data mining dapat diartikan sebagai mengekstraksi atau menggali pengetahuan dari data yang berjumlah besar.

Pada dasarnya data mining berhubungan erat dengan analisis data, dan penggunaan perangkat lunak untuk mencari pola dan kesamaan dalam sekumpulan data. Ide dasarnya adalah menggali sumber yang berharga dari suatu tempat yang sama sekali tidak diduga, seperti perangkat lunak data mining mengekstraksi pola yang sebelumnya tidak terlihat atau tidak begitu jelas sehingga tidak seorang pun yang memperhatikan sebelumnya. Analisa data mining berjalan pada data yang cenderung terus membesar dan teknik terbaik yang digunakan kemudian berorientasi kepada data berukuran sangat besar untuk mendapatkan kesimpulan dan keputusan paling layak [1]. Berikut merupakan gambaran tahapan data mining yang ditunjukkan pada Gambar 1.

Tahapan yang dipresentasikan dalam Gambar 1: mengilustrasikan bagaimana tiap proses bersifat interaktif dimana pemakaian terlibat langsung

- Target data
- pra pemrosesan
- transformasi data
- teknik data mining
- evaluasi pola.

B. Algoritma *K-NN*

Algoritma *K-NN* adalah suatu metode yang menggunakan algoritma supervised, Tujuan dari algoritma *K-NN* adalah untuk mengklasifikasi objek baru berdasarkan atribut dan data sampel. Hasil dari data uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *K-NN*.

Prinsip kerja *K-Nearest Neighbor* yaitu mencari jarak antara dua titik yaitu titik pelatihan dan titik uji, yang kemudian dilakukan *evaluasi* dengan *K* tetangga terdekatnya dalam data pelatihan. Persamaan perhitungan untuk mencari jarak dihitung berdasarkan jarak *Euclidean*, *Correlation*, *Cosine* dan *Cityblock*. Pada penelitian ini parameter yang digunakan adalah *Euclidean*. Rumus menghitung jarak *Euclidean* sebagai berikut : dengan menggunakan rumus *euclidean* [5].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Dimana, *d* adalah jarak antara titik pada data pelatihan *x* dan titik data uji *y* yang akan diklasifikasikan, dimana $x = x_1, x_2, \dots, x_i$ dan $y = y_1, y_2, \dots, y_i$ dan merepresentasikan nilai dari atribut serta *n* merupakan dimensi data atribut [5]. Langkah-langkah untuk menghitung algoritma *K-NN*

- Menentukan parameter *K* (Jumlah tetangga terdekat)
- menghitung kuadrat jarak *Euclid* masing-masing objek terhadap data sampel yang diberikan.
- kemudian mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclid* terkecil
- mengumpulkan kategori *Y* (Klasifikasi *Nearest Neighbor*)

- e) dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksi nilai *queri istance* yang telah dihitung.

C. Pengukuran Kinerja Klasifikasi

Pengujian kinerja sistem klasifikasi pada algoritma *K-NN* ini dapat dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* ini alat yang berguna untuk menganalisis seberapa baiknya klasifikasi yang kita pakai dapat mengenali pola dari kelas yang berbeda seperti ditunjukkan pada *Ttable 1*.

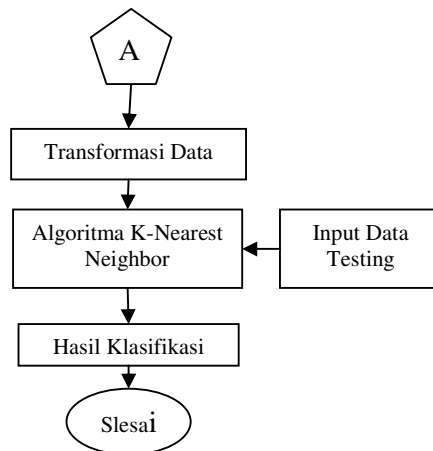
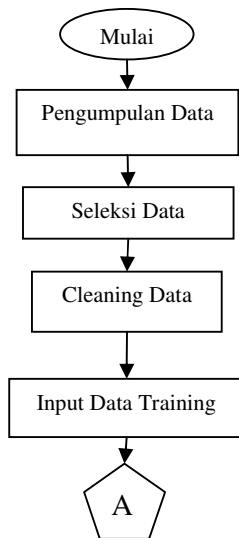
TABLE I
CONFUSION MATRIX

f_{ij}		Kelas Prediksi	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas=1	f_{11}	f_{10}
	Kelas = 0	f_{01}	f_{00}

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{jumlah data diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \\
 &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (2)
 \end{aligned}$$

III. ANALISIS TAHAPAN PENELITIAN

Pada Tahap ini akan dilakukan analisa bagaimana penerapan algoritma *K-NN* menggunakan rumus *euclidean distance* untuk menyelesaikan masalah berupa membagi data sampel (*training*) dan data uji (*testing*), kemudian data uji diklasifikasi menggunakan *k-nearest neighbor*. Berikut ilustrasi penggambaran model Klasifikasi secara umum, dapat dilihat pada Gambar 2.

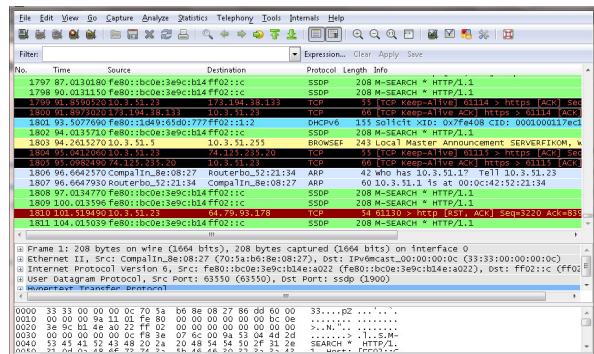


Gambar 2. Metodologi Penelitian

Berdasarkan tahapan pada Gambar2, maka dapat dijabarkan secara lebih terperinci sebagai berikut :

A. Teknik Pengambilan Data Trafik Internet

Pengambilan trafik jaringan menggunakan aplikasi *wireshark*, menghasilkan kurang lebih hingga puluhan juta *record* Trafik Jaringan. Model capture trafik jaringan dari *wireshark* adalah seperti pada Gambar 3.



Gambar 3. Trafik Jaringan

B. Seleksi Data

Pada tahap ini dilakukan dengan memilih data yang akan digunakan sebagai proses penghitungan klasifikasi. Format *record* dari hasil *network capture* yaitu berupa data mentah akan diolah dan difilter menggunakan aplikasi *Pentaho*. *Counting* dilakukan untuk menghitung *record traffic* yang sama. Pada data filter, *ffield* yang dihitung nanti menggunakan metode *K-NN* hanya menggunakan field *Protocol*, *Length* dan jumlah *counting*. sehingga dapat dijadikan acuan dalam penentuan *QoS* dengan menganalisa jumlah data *counting*.

C. Klasifikasi dengan Algoritma K-NN

Pada tahapan pembentukan klasifikasi menggunakan *K-NN* adalah dengan menghitung jarak semua *variabel* pada data uji dengan semua data sampel menggunakan rumus *euclidean distance*. Selanjutnya setelah nilai jarak *euclidean* pada data



sampel diketahui, maka dapat ditentukan *cluster-cluster* pada data *sampel* berdasarkan kedekatan nilai-nilai yang dihasilkan dari perhitungan jarak *Euclidean* tersebut. Pembentukan klasifikasi akan dilakukan pada pola k berbeda yaitu k=3, k=5, k=7 dan k=9.

Data trafik jaringan yang dipetakan dalam metode *K-NN* sebagai kelas klasifikasi yang ditunjukkan pada Tabel II sebagai berikut :

TABEL II
DATA TRAFIK JARINGAN

Jenis Protokol	length Range	counting Range	Prioritas
ARP, DHCP, DHCPv6, HTTP, DNS, TCP, ICMP, ICMPv6, IGMP, MNDP, MDNS, , NBNS, NTP, SSDP, SSHv2	0 – 32, 33 – 64, 65 – 128, 129 – 256, 257 – 512, 513 – 1024, 1025 – 2048	0 – 500, 501 – 1000, 1501 – 2000, 2001 – 2500, 2501 – 3000, x > 3000	Rendah, Menengah Tinggi

Data yang digunakan dalam penelitian ini yaitu sebanyak 6256 data hasil klasifikasi dengan atribut *protocol*, *length*, *counting*, *length range* dan *counting range* dan prioritas. Untuk melakukan inisialisasi, penelitian ini mengusulkan model untuk menentukan kelas label seperti ditunjukkan pada Tabel III, yaitu sebagai berikut.

TABEL III
MODEL KLAS LABEL

Range	Prioritas		
	Rendah	Menengah	Tinggi
Length	33 – 64	129 – 256	513 - 1024
	65 - 128	257 - 512	1025 - 2048
Counting	0 - 500	1501 –	2501 - 3000 x > 3000
	501 - 1000	2000	
	1001 - 1500	2001 –	
		2500	

Seperti ditunjukkan pada Tabel III, rentang batas minimum dan maksimum setiap *prioritas* didasarkan pada hasil data trafik jaringan. Dari aturan model tersebut penelitian ini mengajukan *inisialisasi* label kelas dalam pembentukan data latih yang dilakukan secara manual, seperti ditunjukkan pada Table IV sebagai berikut:

TABEL IV
INISIALISASI KELAS LABEL

Protokol	Length	Counting	Length Range	Counting Range	Prioritas
ARP	118	6095	65 - 128	X > 3000	Rendah
ARP	128	1743	65 - 128	1501 - 2000	Menengah
ARP	136	8	129 - 256	0 - 500	Menengah
DHCP	301	3	257 - 512	0 - 500	Menengah
DHCP	320	2	257 - 512	0 - 500	Menengah
DHCP	331	1	257 - 512	0 - 500	Menengah
DHCP	342	616	257 - 512	501 - 1000	Menengah
DHCP	344	51	257 - 512	0 - 500	Menengah
DHCP	346	225	257 - 512	0 - 500	Menengah
DHCP	347	3	257 - 512	0 - 500	Menengah
DHCP	348	6	257 - 512	0 - 500	Menengah
DHCP	349	1	257 - 512	0 - 500	Menengah
DHCP	350	2	257 - 512	0 - 500	Menengah
DHCP	351	3	257 - 512	0 - 500	Menengah
DHCP	352	5	257 - 512	0 - 500	Menengah
DHCP	353	1	257 - 512	0 - 500	Menengah
DHCP	354	80	257 - 512	0 - 500	Menengah
DHCP	355	7	257 - 512	0 - 500	Menengah
DNS	73	2754	65 - 128	2501 - 3000	Tinggi
DNS	74	3577	65 - 128	X > 3000	Tinggi
DNS	75	3950	65 - 128	X > 3000	Tinggi
DNS	76	5100	65 - 128	X > 3000	Tinggi

IV. HASIL DAN PEMBAHASAN

A. Evaluasi Mekanisme Penelitian

Berikut adalah langkah-langkah perhitungan secara manual atau menentukan prioritas suatu trafik jaringan menggunakan Algoritma K-NN.

- Menghitung jarak total semua parameter pada data *training* dan data *testing* dengan menggunakan persamaan (1). Untuk perhitungan beberapa parameter yang digunakan adalah .
 - Data Sampel yang digunakan sebanyak 5 buah, yang ditunjukkan pada Tabel V.
 - Data Testing dengan data parameter seperti ditunjukkan pada Tabel VI.

Selanjutnya mencari perbandingan jarak terpendek dengan menggunakan persamaan 1, maka dapat dipaparkan sebagai berikut.

$$D(1) = \sqrt{(136 - 346)^2 + (1743 - 225)^2} = \sqrt{46.404} = 6.81$$

$$D(2) = \sqrt{(320 - 346)^2 + (2 - 225)^2} = \sqrt{725.729} = 26.93$$

$$D(3) = \sqrt{(1328 - 346)^2 + (4 - 225)^2} = \sqrt{1.013.155} = 31.83$$

$$D(4) = \sqrt{(333 - 346)^2 + (123 - 225)^2} = \sqrt{10.573} = 3.25$$

$$D(5) = \sqrt{(340 - 346)^2 + (196 - 225)^2} = \sqrt{877} = 29.61$$

Maka didapatkan hasil perbandingan jarak terpendek seperti ditunjukkan pada Tabel VII.

- Selanjutnya mengurutkan hasil perhitungan jarak yang diperoleh pada langkah 1 yang mana data tersebut diurutkan berdasarkan jarak terkecil sampai jarak terbesar, seperti yang ditunjukkan pada Tabel VIII.

TABEL V
DATA TESTING (UJI)

No id	Protokol	Lenght	Counting	Length Range	Counting Range	Prioritas
1	ARP	136	1743	65 – 128	1501- 2000	Menengah
2	DHCP	320	2	257 – 512	0 – 500	Menengah
3	HTTP	1328	4	1025 – 204	0 – 500	Rendah
4	DNS	333	123	257 – 512	0 – 500	Menengah
5	DNS	340	196	257 – 512	0-500	Menengah

TABEL VI
DATA SAMPEL (TRAINING)

No Id	Protokol	Length	Counting	Prioritas
1	DNS	346	225	?

TABEL VII
DATA HASIL PERHITUNGAN PERBANDINGAN JARAK

No id	Protokol	Lenght	Counting	Length Range	Counting Range	Prioritas	Jarak
1	ARP	136	1743	65 – 128	1501 – 2000	Menengah	6.81
2	DHCP	320	2	257 – 512	0 – 500	Menengah	26.93
3	HTTP	1328	4	1025 – 204	0 – 500	Rendah	31.83
4	DNS	333	123	257 – 512	0 - 500	Menengah	3.25
5	DNS	340	196	257 – 512	0-500	Menengah	29.61

TABEL VIII
DATA YANG TELAH DIURUTKAN

No id	Protokol	Lenght	Counting	Length Range	Counting Range	Prioritas	Jarak
4	DNS	333	123	257- 512	0 - 500	Menengah	3.25
1	ARP	136	1743	65 - 128	1501-2000	Menengah	6.81
2	DHCP	320	2	257- 512	0 - 500	Menengah	26.93
5	DNS	340	196	257- 512	0-500	Menengah	29.61
3	HTTP	1328	4	1025-204	0 - 500	Rendah	31.83



3. Menentukan ketetangaan terdekat untuk mendapatkan prioritas yang cocok dengan data *testing*.

Berdasarkan hasil perhitungan jarak ditetapkan $k = 5$, Maka didapatkan hasil : Menengah = 4, Rendah = 1, Maka bisa ditetapkan bahwa, sampel trafik jaringan yaitu dengan *length* 346 dan *counting* 225 menghasilkan prioritas Menengah. Seperti ditunjukkan pada Tabel IX.

B. Hasil Klasifikasi Algoritma K-NN

Berdasarkan hasil pengujian klasifikasi trafik jaringan menggunakan *algoritma K-NN* pada MATLAB, yaitu menggunakan 6257 sebagai data latih dan 5596 sebagai data uji maka didapatkan hasil seperti ditunjukkan pada Tabel X.

Tabel X. menunjukan hasil pengujian yang dilakukan terhadap data baru maka dihasilkan data hasil klasifikasi menggunakan algoritma *K-NN* menggunakan MATLAB. Bahwa protokol ARP dengan *length* 60 dan *counting* 552581 menghasilkan klasifikasi dengan prioritas tinggi.

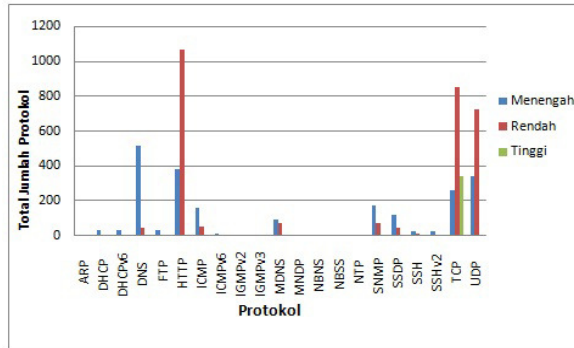
Hal ini disebabkan karena pengguna internet yang menggunakan protokol ARP lebih sering dengan ukuran *length* 60 byte oleh karena itu, protokol ARP akan diberikan prioritas tinggi. Dengan diberikan prioritas tinggi kepada pengguna internet maka pengguna internet merasa nyaman menggunakan protokol komunikasi ARP tersebut.

TABEL IX
DATA HASIL KLASIFIKASI

No Id	Protokol	Length	Counting	Prioritas
1	DNS	346	225	Menengah

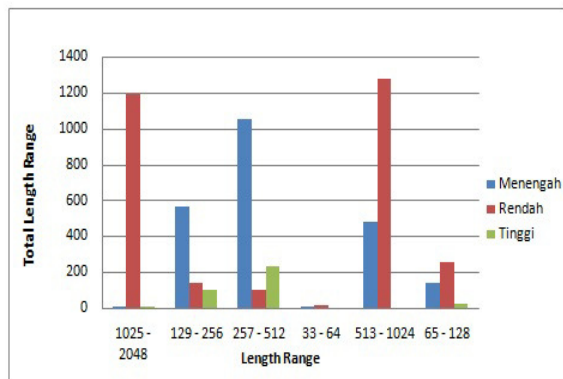
TABEL X
Hasil Klasifikasi Trafik Jaringan menggunakan Algoritma K-NN

No	Protokol	Length	Counting	Prioritas
2	ARP	60	552581	Tinggi
3	ARP	110	96	Rendah
4	ARP	118	1920	Menengah
5	ARP	128	1058	Rendah
6	DHCP	342	249	Menengah
7	DHCP	343	1	Menengah
8	DHCP	344	27	Menengah
9	DHCP	346	93	Menengah
10	DHCP	347	5	Menengah
11	DHCP	348	4	Menengah
12	DHCP	350	3	Menengah
13	DHCP	351	1	Menengah
14	DHCP	353	3	Menengah
15	DHCP	354	37	Menengah
16	DHCP	355	1	Menengah
17	DHCP	356	1	Menengah
18	DHCP	357	3	Menengah
19	DHCP	360	2	Menengah
20	DHCP	362	7	Menengah
21	DHCP	365	1	Menengah
22	DHCP	370	4	Menengah
23	DHCP	386	1	Menengah
24	DHCP	414	104	Menengah
25	DHCP	418	104	Menengah
26	DHCP	420	2	Menengah
27	DHCP	421	2	Menengah
28	DHCP	422	2	Menengah
29	DHCP	423	2	Menengah
30	DHCP	424	8	Menengah
31	DHCP	427	12	Menengah
32	DHCP	428	4	Menengah
33	DHCP	429	6	Menengah
34	DHCP	430	7	Menengah



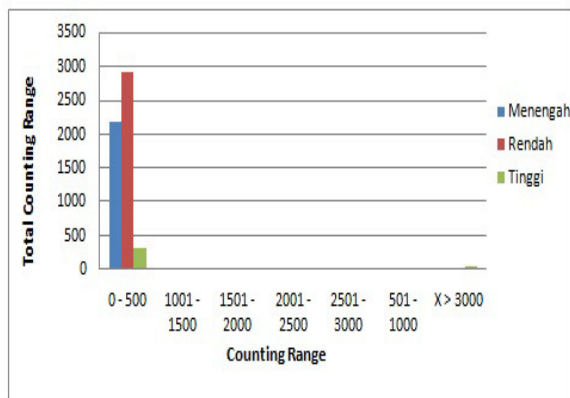
Gambar 4. Banyak protokol berdasarkan prioritas

Dari Gambar 4, dapat disimpulkan bahwa hasil klasifikasi trafik jaringan dari sisi protokol terhadap prioritas menjelaskan bahwa protokol HTTP memiliki tingkat kemunculan yang tinggi maka prioritas yang diberikan oleh sistem adalah prioritas rendah.



Gambar 5. Banyak length range berdasarkan prioritas

Pada Gambar 5, menunjukkan bahwa length yang diberikan prioritas rendah yaitu terhadap komunikasi yang memiliki ukuran data dari 1025–2048 byte dan 513 – 1024 byte.



Gambar 6. Banyak Counting Range berdasarkan prioritas

I Komang Kompyang Agus Subrata : Klasifikasi Penggunaan Protokol.....

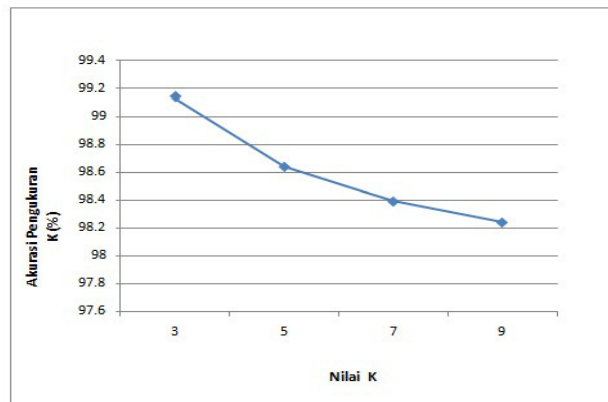
Dari Gambar 6, dapat disimpulkan bahwa counting range 0 – 500 memiliki tingkat kemunculan (banyaknya data record aktifitas yang sama) yang tinggi maka, akan diberikan prioritas rendah pada count range 0 – 500.

C. Hasil Akurasi

Pada pengujian ini didapatkan akurasi tertinggi pada K-NN mencapai 99.14% seperti yang ditunjukkan pada Tabel XI.

TABEL XI Hasil Pengitungana akurasi klasifikasi K-NN

Data Uji	Nilai K	Akurasi
6257	3	0,9914
	5	0,9864
	7	0,9839
	9	0,9824



Gambar 5. Pengaruh Nilai k pada 6257 Data Latih

Kesimpulan yang diperoleh yaitu, dari 6257 data uji (testing) kemudian dilakukan pengukuran akurasi dengan menggunakan parameter k yang berbeda-beda, maka nilai akurasi yang didapat berbeda-beda, tetapi memiliki nilai akurasi yang hampir sama sehingga didapatkan tingkat akurasi kecocokan tertinggi pada hasil sebenarnya terhadap hasil dari prediksi dengan menggunakan algoritma K-Nearest Neighbor didapatkan pada nilai k = 3, yaitu sebesar 99,14 %. Dari grafik diketahui bahwa semakin tinggi jumlah k, maka semakin rendah akurasi yang didapatkan. Hal ini dikarenakan rentang kelas pada k yang semakin banyak memberikan nilai sensitifitas yang besar pada penentuan prediksi.

V. KESIMPULAN

Peper ini telah memaparkan sebuah model penentuan klasifikasi trafik internet menggunakan algoritma K-NN dengan inialisasi yang telah ditentukan. Informasi data trafik internet Universitas Udayana diambil atau diperoleh melalui mekanisme Capture data menggunakan aplikasi perangkat lunak wireshar. Hasil data trafik capture akan diolah dengan



proses data mining dengan menggunakan algoritma K-NN. Algoritma K-NN mengklasifikasi QoS berdasarkan tingkat kemiripan data uji dengan data pelatihan.

Berdasarkan analisis data yang telah dilakukan didapatkan sebuah hasil bahwa nilai yang dihasilkan oleh klasifikasi K-NN memiliki tingkat keakuratan yang sangat tinggi. Hal ini dibuktikan dari hasil observasi terhadap dataset trafik jaringan UPT PUSKOM Universitas Udayana dan melalui proses perhitungan menggunakan Algoritma K-NN dengan label kelas dan atribut yang telah dipaparkan di pembahasan sebelumnya, didapatkan sebuah hasil bahwa nilai yang dihasilkan oleh klasifikasi K-NN memiliki tingkat keakuratan yang sangat tinggi. Hal ini dibuktikan dengan hasil perhitungan yang mencapai nilai 99,14 % yaitu dengan perhitungan $k = 3$.

Besar nilai k dan banyak data *training* yang digunakan dalam melakukan klasifikasi sangat berpengaruh dalam menentukan hasil klasifikasi. Semakin besar nilai k yang digunakan, maka akan menghasilkan klasifikasi yang tidak jelas dan juga akan menghasilkan kesalahan klasifikasi yang besar.

REFERENSI

- [1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, 2012.
- [2] J. Zhang, C. Chen, Y. Xiang, W. Zhou, and Y. Xiang, "Internet Traffic Classification by Aggregating Correlated Naive Bayes Predictions," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 5–15, Jan. 2013.
- [3] M. Sudarma and D. P. Hostiadi, "Klasifikasi Penggunaan Protokol Komunikasi Pada Network Traffic Menggunakan Naive Bayes Sebagai Penentuan QoS," *Pros. CSGTEIS 2013*, 2013.
- [4] "Klasifikasi Data Minuman Wine Menggunakan Algoritma K-Nearest Neighbor." [Online]. Available: <https://www.scribd.com/document/317831440/Klasifikasi-Data-Minuman-Wine-Menggunakan-Algoritma-K-Nearest-Neighbor>. [Accessed: 12-Jul-2016].
- [5] W. Hidayat, E. M. Dharma, and M. A. Bijaksana, "PENERAPAN K-NEAREST NEIGHBOUR UNTUK KLASIFIKASI GAMBAR LANDSCAPE BERDASARKAN FITUR WARNA DAN TEKSTUR," 2005.