INTERNET NETWORK CLASSIFICATION IN MALIKUSSALEH UNIVERSITY USING NAÏVE BAYES METHOD

Eva Darnila[✉]**, Zara Yunizar, Dhyra Gibran Alinda**
Departement of Informatics, Universitas Malikussaleh, Aceh Utara, Indonesia
Email: eva.darnila@unimal.ac.id

DOI: https://doi.org/10.46880/jmika.Vol5No1.pp48-53

ABSTRACT

The utilize of web systems at this time is exceptionally critical, particularly for the world of instruction. In expansion to the significance of the web arrange, issues frequently emerge on the web arrange due to an expansive number of clients, the issues can be gotten at the UNIMAL campus incorporate moderate, harmed, and indeed not sent information to its goal since the organize activity isn't ideal. To be able to optimize the web organize by prioritizing organize activity. In this consider, the Credulous Bayes calculation is utilized for the classification handle of organize activity capture information. The application utilized to capture organize activity is the Wireshark application. Utilizing the Credulous Bayes calculation to watch the comes about of organize test information through a calculation prepare that has tall precision. To be specific at the Workforce of Designing 95.73% with a likelihood of testing comes about is 0.00015946 web browsing, 0.00000007 downloads, 0.00008691 gushing, 0.00008497 social media, 0.00000014 floodings.

Keyword: Wireshark, Classification, Naïve Bayes Methods.

INTRODUCTION

In today's advanced life, the improvement of data innovation is so quick that it is taken after by the improvement of communication, particularly the improvement of the web. This data innovation has made the communication preparation less demanding, specifically by dispensing with remove and time which were considered as limitations, in this way requiring a incredible require for Indonesia to utilize the Web. The utilize of the web is a critical prerequisite to bolster campus execution and exercises. The foremost imperative portion of the Web framework given by the campus is the accessibility of adequate transmission capacity to guarantee smooth information activity on the Web. More often than not, the sum of accessible transfer speed is inadequately, particularly amid top hours and dynamic college periods (Prathivi, 2015).

With the tall request and utilization of arrange innovation, clients anticipate the arrange to attain greatest levels of efficiency and security. For illustration, there's a problem at the UNIMAL campus, since arrange activity isn't ideal, the information sent is exceptionally moderate, harmed, and indeed cannot reach its goal. Subsequently, analyzing arrange activity is one way to get it the utilize of organize communication conventions, so that it can be utilized as a premise for deciding arrange activity need (Dibawan, Widyantara, & Linawati, 2016). The rectify classification of Web organize activity is exceptionally vital, particularly in arrange design plan, organize administration, and arrange security. The classification is based on the number of sorts of communication exercises controlled by organize conventions. Analyzing organize activity is one way to get it to utilize arrange communication conventions so that it can be utilized as a premise for deciding arrange activity need (Subrata, Widyantara, & Linawati, 2016).

With respect to the classification of web activity, a few ponders have been carried out utilizing the Gullible Bayes calculation. As can be seen from the clarification over, the application of the Gullible Bayes calculation may be a solution that can be connected to the web activity classification instrument. In expansion, analysts will take Malikussaleh College Web activity as a case to assist assess the application performance of the Credulous Bayes calculation within the campus Web activity classification. The objective is to progress execution on the UNIMAL arrange and as a reference for directors in organize administration. The classification instrument utilized is to decide the likelihood of UNIMAL web activity being captured on a UNIMAL organize switch by utilizing the Wireshark application.

RESEARCH METHODS

Classification is the method of finding a demonstrate or work that portrays or classifies information into classes. Classification includes the method of analyzing the characteristics of a question and inserting that object into a class. In common, classification may be a demonstrate made to portray a foreordained set of information classes or concepts. Construct models by analyzing speculative records into foreordained classes (called course attributes) (Dibawan et al., 2016).

The Naive Bayes algorithm is a calculation that's included within the classification procedure. Naïve Bayes could be a classification proposed by the British researcher Thomas Bayes utilizing likelihood and measurable strategies that foresee future openings based on past encounters, so it is called the Bayes hypothesis. The Bayes hypothesis condition is as takes after (Dibawan et al., 2016):

$$P(H|X) = \frac{P(X|H) * P(H)}{P(X)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where:
X : Data with an unknown class.
H : Data hypothesis X a class.
P(H│X) : The probability of hypothesis H based on condition X (Posterior).
P(H) : Hypothesis probability H (Prior).
P(X│H) : Probability X based on the conditions in the hypothesis
P(X) : Probability X.

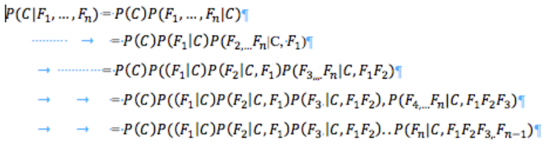
To clarify the Credulous Bayes hypothesis, it ought to be famous that the classification preparation requires a part of clues to decide which category is appropriate for the test being analyzed. In this manner, the Bayes hypothesis specified over is balanced as takes after:

$$P(H|X) = \frac{P(C)P(F_1 \ldots F_n|C)}{P(F_1 \ldots F_n)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where the variable C speaks to the course, whereas the variable F1 ... Fn speaks to the characteristics of the enlightening required to perform the classification. At that point, the equation depicts the likelihood of a particular characteristic test entering course C (back), to be specific the chance of course C (sometime recently the passage of the test, it is frequently called a earlier), increase by the likelihood of event of the characteristic test C (probability), separated by the likelihood of event of the characteristics of the test at the same time. worldwide (prove), hence, the over equation can too be composed essentially as takes after

$$Posterior = \frac{likelihood\ x\ prior}{evidence} \quad \ldots\ldots\ldots\ldots\ldots (3)$$

Evidence value is always assigned to each category in the sample, and the posterior value is then compared with the posterior value of other categories to determine the sample category, further elaboration of the Bayes formula is done by elaboration (C | F1, .., Fn) Use the following multiplication rule:



$$P(C|F_1, \ldots, F_n) = P(C)P(F_1, \ldots, F_n|C)$$
$$= P(C)P(F_1|C)P(F_{2,\ldots}F_n|C, F_1)$$
$$= P(C)P((F_1|C)P(F_2|C, F_1)P(F_{3,\ldots}F_n|C, F_1F_2)$$
$$= P(C)P((F_1|C)P(F_2|C, F_1)P(F_3|C, F_1F_2), P(F_{4,\ldots}F_n|C, F_1F_2F_3)$$
$$= P(C)P((F_1|C)P(F_2|C, F_1)P(F_3|C, F_1F_2)..P(F_n|C, F_1F_2F_{3,.}F_{n-1})$$

It appears that the results of these descriptions lead to increasingly complex determinants affecting the probability value, which are almost impossible to analyze individually, and as a result of which these calculations become difficult to carry out, it is here that the assumption of very high independence (naivety) is used.). Whereas each clue (F_1 〖, F〗 _2 〖, ... F〗 _n) is independent (independent) of each other, with this assumption, the following similarities apply as follows:

$$P(P_i|P_j) = \frac{P(P_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_i) = P(F_i|C) \ldots\ldots\ldots (4)$$

From the above equation, it can be concluded that the naive assumption of independence makes the conditions of probability simple, allowing the calculation to be carried out, and then, the translation of P (C | F_1, ..., F_n) can be simplified to:

$$P(C|F_1, \ldots, F_n) = P(C)P(F_1|C)P(F_2|C)P(F_3|C) \ldots = (C) \prod_i^n = 1\ P(F_i|C) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

The last stage is to test the accuracy of data mining using the Naive Bayes algorithm. The method of calculating the accuracy is as follows.

The formula for calculating insurance is written as below:

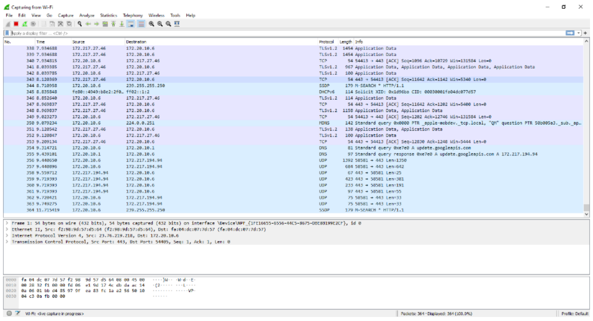
a. *Acuracy* useful for measuring the performance of a method.

$$Acuracy = \frac{\text{the amount of data predicted is correct}}{\text{the total number of predictions}}\ x\ 100\% \ldots. (6)$$

b. *Error is useful for measuring the degree of mismatch.*

$$Error = \frac{\text{the amount of data predicted is wrong}}{\text{the total number of predictions}} \; x \; 100\% ...(7)$$

Retrieving Network Traffic using the Wireshark application is done by capturing traffic on the network. In recorded network traffic, about hundreds of thousands of traffic are generated every five minutes. However, the number of records generated is not the same. The unequal number of traffic records is caused by the inequality of the communication model in the computer network carried out by the user. Network traffic capture model from wireshark is as shown below:


Capturing from Wi-Fi
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
No. Time Source Destination Protocol Length Info

```
338 7.934688  172.217.27.46          172.20.10.6      TLSv1.2  1454 Application Data
339 7.934688  172.217.27.46          172.20.10.6      TLSv1.2  1454 Application Data
340 7.934815  172.20.10.6            172.217.27.46    TCP        54 54413 → 443 [ACK] Seq=1096 Ack=10729 Win=131584 Len=0
341 8.039385  172.217.27.46          172.20.10.6      TLSv1.2   967 Application Data, Application Data, Application Data, Application Data
342 8.039785  172.20.10.6            172.217.27.46    TLSv1.2   100 Application Data
343 8.120369  172.217.27.46          172.20.10.6      TCP        54 443 → 54413 [ACK] Seq=11642 Ack=1142 Win=5340 Len=0
344 8.710958  172.20.10.6            239.255.255.250  SSDP      179 M-SEARCH * HTTP/1.1
345 8.835848  fe80::4549:b8e2:2f0…   ff02::1:2        DHCPv6    114 Solicit XID: 0x1d96ba CID: 00030001fa04dc077d57
346 8.852640  172.20.10.6            172.217.27.46    TLSv1.2   114 Application Data
347 8.969837  172.217.27.46          172.20.10.6      TCP        54 443 → 54413 [ACK] Seq=11642 Ack=1202 Win=5400 Len=0
348 8.969837  172.217.27.46          172.20.10.6      TLSv1.2  1158 Application Data, Application Data
349 9.023273  172.20.10.6            172.217.27.46    TCP        54 54413 → 443 [ACK] Seq=1202 Ack=12746 Win=131584 Len=0
350 9.079234  172.20.10.6            224.0.0.251      MDNS      142 Standard query 0x0000 PTR _apple-mobdev._tcp.local, "QM" question PTR 50b005a3._sub._ap…
351 9.120542  172.217.27.46          172.20.10.6      TLSv1.2   138 Application Data, Application Data
352 9.120847  172.20.10.6            172.217.27.46    TLSv1.2   100 Application Data
353 9.209134  172.217.27.46          172.20.10.6      TCP        54 443 → 54413 [ACK] Seq=12830 Ack=1248 Win=5444 Len=0
354 9.314721  172.20.10.6            172.20.10.1      DNS        81 Standard query 0xe7e0 A update.googleapis.com
355 9.439101  172.20.10.1            172.20.10.6      DNS        97 Standard query response 0xe7e0 A update.googleapis.com A 172.217.194.94
356 9.440650  172.20.10.6            172.217.194.94   UDP      1392 58581 → 443 Len=1350
357 9.440896  172.20.10.6            172.217.194.94   UDP       684 58581 → 443 Len=642
358 9.559712  172.217.194.94         172.20.10.6      UDP        67 443 → 58581 Len=25
359 9.719393  172.217.194.94         172.20.10.6      UDP       423 443 → 58581 Len=381
360 9.719393  172.217.194.94         172.20.10.6      UDP       233 443 → 58581 Len=191
361 9.719393  172.217.194.94         172.20.10.6      UDP        97 443 → 58581 Len=55
362 9.720421  172.20.10.6            172.217.194.94   UDP        75 58581 → 443 Len=33
363 9.749275  172.20.10.6            172.217.194.94   UDP        75 58581 → 443 Len=33
364 11.715419 172.20.10.6            239.255.255.250  SSDP      179 M-SEARCH * HTTP/1.1

> Frame 1: 54 bytes on wire (432 bits), 54 bytes captured (432 bits) on interface \Device\NPF_{1FE16655-6556-44C5-8675-D0E88199E2CF}, id 0
> Ethernet II, Src: f2:98:9d:57:d5:64 (f2:98:9d:57:d5:64), Dst: fa:04:dc:07:7d:57 (fa:04:dc:07:7d:57)
> Internet Protocol Version 4, Src: 23.76.219.218, Dst: 172.20.10.6
> Transmission Control Protocol, Src Port: 443, Dst Port: 54405, Seq: 1, Ack: 1, Len: 0

0000  fa 04 dc 07 7d 57 f2 98  9d 57 d5 64 08 00 45 00   ····}W·· ·W·d··E·
0010  00 28 32 f1 00 00 fd 06  e1 9d 17 4c db da ac 14   ·(2····· ···L····
0020  0a 06 01 bb d4 85 97 9f  ea 83 fc 1a a2 56 50 10   ········ ·····VP·
0030  04 c3 0a fb 00 00                                  ······

Wi-Fi: <live capture in progress>    Packets: 364 · Displayed: 364 (100.0%)    Profile: Default
```

Figure 1. Capturing Network traffic

Wireshark is one of the numerous Arrange Analyzer disobedient that are broadly utilized by Arrange Chairmen to analyze arrange execution. Wireshark is broadly enjoyed since of its interface that employments a Graphical Client Interface (GUI) or a graphical show. Wireshark is utilized for arranging investigating, examination, program and communication convention improvement, and instruction. Wireshark is broadly utilized by Arrange Directors to analyze arrange execution. Wireshark is able to capture information/data passing through an organization that we watch within the frame of organized activity. The benefits of utilizing the Wireshark application areas takes after (Sudarma & Hostiadi, 2013), Capturing data or bundle information sent and gotten on computer systems, understanding exercises that happen on computer systems, understanding and analyzing the execution of our computer systems, such as information get to / sharing speed and a arrange association to the web, watching the security of the computer network.

The plot of the UNIMAL web organize classification framework utilizing the Naïve Bayes calculation is as takes after:

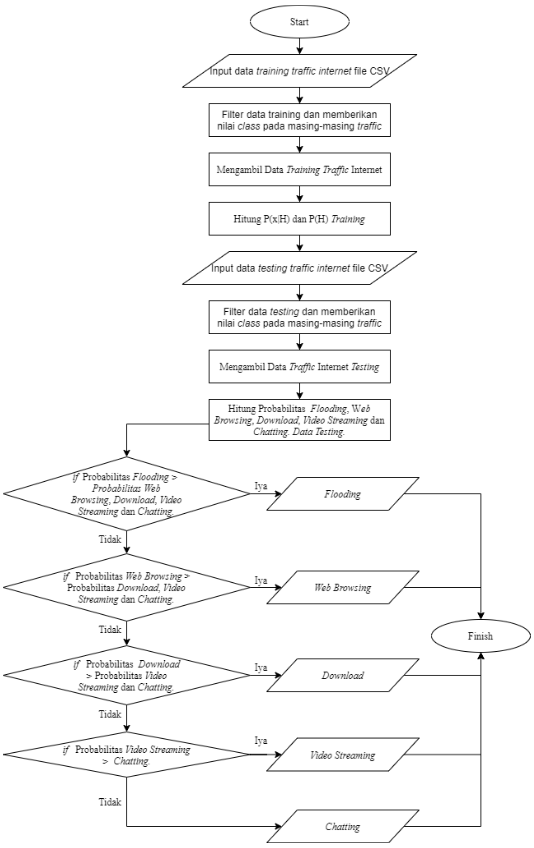

Start
Input data *training traffic internet* file CSV
Filter data training dan memberikan nilai *class* pada masing-masing *traffic*
Mengambil Data *Training Traffic* Internet
Hitung P(x|H) dan P(H) *Training*
Input data *testing traffic internet* file CSV
Filter data *testing* dan memberikan nilai *class* pada masing-masing *traffic*
Mengambil Data *Traffic* Internet *Testing*
Hitung Probabilitas *Flooding*, *Web Browsing*, *Download*, *Video Streaming* dan *Chatting*. *Data Testing*.
if Probabilitas *Flooding* > *Probabilitas Web Browsing*, *Download*, *Video Streaming* dan *Chatting*.
Iya
Flooding
Tidak
if Probabilitas *Web Browsing* > Probabilitas *Download*, *Video Streaming* dan *Chatting*.
Iya
Web Browsing
Tidak
if Probabilitas *Download* > Probabilitas *Video Streaming* dan *Chatting*.
Iya
Download
Tidak
if Probabilitas *Video Streaming* > *Chatting*.
Iya
Video Streaming
Tidak
Chatting
Finish

Figure 2. System Scheme

RESULT AND DISCUSSION

Building workforce preparing information were collected from 1/8/2020 - 12/9/2020, with a add up to web activity information of 10,945, comprising of 9,623 web browsing activity information, 378 spilling activity information, 387 social media information, 66 download information, and 491 identified information as flooding. Testing information or test information could be a raw fabric that's intentionally made to assist the investigation process in a information mining framework, in more detail, testing information can be seen within the taking after table:

Table 1. Internet traffic data table

NO	DOMAIN	PROTOCOL	PORT	CLASSIFICATION
1	external.fplm4-1.fna.fbcdn.net	HTTPS	443	?

For data testing at the Engineering faculty, it can be calculated P (Ci) or the number of classes / labels. The number of classes / labels is 5, namely web browsing, downloading, streaming, flooding and social media.

P (Y = label / class) = web browsing, number of labels, class / total amount of data
P (Y = web browsing) = 9623/10945 = 0.87921425
P (Y = download) = 66/10945 = 0.00603015
(Y = streaming) = 378/10945 = 0.03453632
P (Y = social media) = 387/10945 = 0.03535861
P (Y = flooding) = 491/10945 = 0.04486067

In calculating P (X | Ci) or the number of the same cases with the same class. To avoid results with the number 0 (zero), laplace correction is used for each calculated data.

In label class Y = web browsing, the amount of data labeled class Y = web browsing is 9623

$$\mathrm{P}(domain = \mathrm{external.fplm4} - 1.\mathrm{fna.fbcdn.net} \mid \mathrm{Y} = web\ browsing) = \frac{(1+1)}{(9623+5)} = 0.00020773$$

$$\mathrm{P}(protocol = \mathrm{HTTPS} \mid \mathrm{Y} = web\ browsing) = \frac{(9036+1)}{(9623+5)} = 0.93861654$$

$$\mathrm{P}(port = 443 \mid \mathrm{Y} = web\ browsing) = \frac{(8955+1)}{(9623+5)} = 0.93020357$$

At Label class Y = download, the amount of data labeled class Y = download is 66

$$\mathrm{P}(domain = \mathrm{external.fplm4} - 1.\mathrm{fna.fbcdn.net} \mid \mathrm{Y} = download) = \frac{(0+1)}{(66+5)} = 0.01408451$$

$$\mathrm{P}(protocol = \mathrm{HTTPS} \mid \mathrm{Y} = download) = \frac{(1+1)}{(66+5)} = 0.02816901$$

$$\mathrm{P}(port = 443 \mid \mathrm{Y} = download) = \frac{(1+1)}{(66+5)} = 0.02816901$$

On Label class Y = streaming, the amount of data labeled class Y = streaming is 378

$$\mathrm{P}(domain = \mathrm{external.fplm4} - 1.\mathrm{fna.fbcdn.net} \mid \mathrm{Y} = streaming) = \frac{(0+1)}{(378+5)} = 0.00261097$$

$$\mathrm{P}(protocol = \mathrm{HTTPS} \mid \mathrm{Y} = streaming) = \frac{(375+1)}{(378+5)} = 0.98172324$$

$$\mathrm{P}(port = 443 \mid \mathrm{Y} = streaming) = \frac{(375+1)}{(378+5)} = 0.98172324$$

In label class Y = social media, the amount of data labeled class Y = social media is 387

$$\mathrm{P}(domain = \mathrm{external.fplm4} - 1.\mathrm{fna.fbcdn.net} \mid \mathrm{Y} = sosial\ media) = \frac{(0+1)}{(387+5)} = 0.00255102$$

$$\mathrm{P}(protocol = \mathrm{HTTPS} \mid \mathrm{Y} = sosial\ media) = \frac{(385+1)}{(387+5)} = 0.98469388$$

$$\mathrm{P}(port = 443 \mid \mathrm{Y} = sosial\ media) = \frac{(374+1)}{(387+5)} = 0.95663265$$

In label class Y = flooding, the amount of data labeled class Y flooding is 491

$$\mathrm{P}(domain = \mathrm{external.fplm4} - 1.\mathrm{fna.fbcdn.net} \mid \mathrm{Y} = flooding) = \frac{(0+1)}{(491+5)} = 0.00201613$$

$$\mathrm{P}(protocol = \mathrm{HTTPS} \mid \mathrm{Y} = flooding) = \frac{(0+1)}{(491+5)} = 0.00201613$$

$$\mathrm{P}(port = 443 \mid \mathrm{Y} = flooding) = \frac{(386+1)}{(491+5)} = 0.78024194$$

To calculate the value of P (Ci) * P (X | Ci) or multiply all the variables / classes of web browsing, downloading, streaming, flooding and social media.

P(Y = web browsing) = P(domain | *web browsing*) * P(protocol | *web browsing*) * P(port | *web browsing*) * P(Y | *web browsing*)
= 0.00020773 * 0.93861654 * 0.93020357 * 0.87921425
= 0.00015946

P(Y = download) = P(domain | *download*) * P(protocol | *download*) * P(port | *download*) * P(Y | *download*)
= 0.01408451 * 0.02816901 * 0.02816901 * 0.00603015
= 0.00000007

P(Y = $streaming$) = P(domain | $streaming$) * P(protocol | $streaming$) * P(port | $streaming$) * P(Y | $streaming$)
= 0.00261097 * 0.98172324 * 0.98172324 * 0.03453632
= 0.00008691

P(Y = $sosial\ media$) = P(domain | $sosial\ media$) * P(protocol | $sosial\ media$) * P(port | $sosial\ media$) * P(Y | $sosial\ media$)
= 0.00255102 * 0.98469388 * 0.95663265 * 0.03535861
= 0.00008497

P(Y = $flooding$) = P(domain | $flooding$) * P(protocol | $flooding$) * P(port | $flooding$) * P(Y | $flooding$)
= 0.00201613 * 0.00201613 * 0.78024194 * 0.04486067
= 0.00000014

Compare the results of the largest probability value showing the status label class web browsing, downloading, streaming, flooding and social media, where the probability of web browsing class is higher with a probability of 0.00015946 so the data above is included in the web browsing class.
The accuracy value in the Engineering Faculty data can be seen in Table 2.

Table 2. Table of accuracy of Engineering faculty data

No	Domain	Protocol	Port	Classification	Prediction
1	external.fplm4-1.fna.fbcdn.net	HTTPS	443	browsing	browsing
2	koinworks.com	HTTPS	443	browsing	browsing
3	v10.events.data.microsoft.com	HTTPS	443	browsing	browsing
4	blogkoinworks.sgp1.digitaloceanspaces.com	HTTPS	443	browsing	browsing
5	icono-49d6.kxcdn.com	HTTPS	443	browsing	browsing
6	dns.msftncsi.com	HTTPS	443	browsing	browsing
7	youtube-ui.l.google.com	HTTPS	443	browsing	browsing
8	s.ytimg.com	HTTPS	443	browsing	browsing
9	d2r1yp2w7bby2u.cloudfront.net	HTTPS	443	browsing	browsing
10	amplify.outbrain.com	HTTPS	443	browsing	browsing
11	wzrkt.com	HTTPS	443	browsing	browsing
12	tr.outbrain.com	HTTPS	443	browsing	browsing
13	api-glb-apse1c.smoot.apple.com	HTTPS	443	browsing	browsing
14	clients1.google.com	HTTPS	443	browsing	browsing
15	a1838.dscb.akamai.net	HTTPS	443	browsing	browsing
16	is-ssl.mzstatic.com.itunes-apple.com.akadns.net	HTTPS	443	browsing	browsing
17	gateway.fe.apple-dns.net	HTTPS	443	browsing	browsing
18	e6858.dsce9.akamaiedge.net	HTTPS	443	browsing	browsing
19	www-cdn.icloud.com.akadns.net	HTTPS	443	browsing	browsing
20	my.in-fbs.com	HTTPS	443	browsing	browsing

$$Acuracy = \frac{the\ number\ of\ correct\ predictions}{the\ total\ number\ of\ predictions}$$
$$= \frac{(211 + 0 + 1 + 4 + 9)}{(220 + 0 + 1 + 4 + 9)}$$
$$= \frac{224}{234} X100\% = 95{,}73\%$$
$$Error = \frac{number\ of\ wrong\ predictions}{the\ total\ number\ of\ predictions}$$
$$= \frac{(10 + 0 + 0 + 0 + 0)}{(221 + 0 + 1 + 4 + 9)}$$
$$= \frac{10}{234} X100\% = 4{,}27\%$$

Accuracy of 95.73%: Classification of internet traffic testing consisting of 220 web browsing, 0 downloads, 4 streams, 1 flooding and 9 social media, accurate 210 web browsing, 0 downloads, 3 streams, 1 flooding and 9 social media, with 95.73% percentage with 243 internet traffic data testing.

Error 4.27%: Classification of internet traffic testing consisting of 220 web browsing, 0 downloads, 4 streaming, 1 flooding and 9 social media, the error is 10 web browsing, 0 downloads, 0 streaming, 0 flooding and 0 social media, with a percentage of 4.27% with 234 internet traffic data testing.

CONCLUSION

The results of the analysis and testing carried out on the previous system, the conclusions drawn, among others, that this application is able to classify internet traffic into flooding, web browsing, downloading, video streaming and social media. The size of the training data affects the length of the classification process, the greater the training data, the longer the classification process.

Training data is taken from internet traffic data from several faculties. At the Faculty of Engineering Data was taken for 30 days, starting from 1/8/2020 - 12/9/2020 a total of 10,945 data, where the results of the probability test data search obtained the following classification 0.00015946 web browsing, 0.00000007 downloads, 0.00008691 streaming, 0.00008497 social media, and 0.00000014 flooding.

At the Faculty of Economics, data is taken for 30 days, starting from 16/9/2020 - 30/10/2020 a total of

3,902 data, where the results of the probability test data search get the following classifications of 0.00022630 web browsing, 0.00000177 downloads, 0.00023456 streaming, 0.00022124 social media, and 0.00000098 flooding.

At the Faculty of Social and Political Sciences, data was taken for 30 days, starting from 2/11/2020 - 14/12/2020 a total of 10,711 data, where the results of the probability test data search were obtained the following classifications were 0.00016715 web browsing, 0.00000078 downloads, 0.00009011 streaming, 0.00008787 social media, and 0.00000016 flooding.

Testing internet traffic classification with the naïve Bayes algorithm uses 234 test data, on 10,945 Faculty of Engineering training data produces 95.73% accuracy and 4.27% error, for 3,902 Faculty of Economics training data produces 48.72% accuracy and 51.28% errors, and 10,711 Faculty of Social and Political Sciences training data resulted in an accuracy of 94.87% and an error of 5.18%.

REFERENCES

Dibawan, I. M. B., Widyantara, I. M. O., & Linawati. (2016). KLASIFIKASI TRAFIK INTERNET KAMPUS BERBASIS PROTOKOL JARINGAN MENGGUNAKAN ALGORITMA NAIVE BAYES. *Jurnal SPEKTRUM*, *3*(2), 7–14.

Prathivi, R. (2015). Klasifikasi Data Trafik Internet Menggunakan Metode Bayes Network (Studi Kasus Jaringan Internet Universitas Semarang). *Jurnal Transformatika*, *12*(2), 42. https://doi.org/10.26623/transformatika.v12i2.81

Subrata, K. K. A., Widyantara, I. M. O., & Linawati, L. (2016). KLASIFIKASI PENGGUNAAN PROTOKOL KOMUNIKASI PADA TRAFIK JARINGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Majalah Ilmiah Teknologi Elektro*, *16*(1), 67. https://doi.org/10.24843/MITE.1601.10

Sudarma, M., & Hostiadi, D. P. (2013). Klasifikasi Penggunaan Protokol Komunikasi Pada Nework Traffic Menggunakan Naive Bayes Sebagai Penentuan QoS. *PROSIDING CSGTEIS 2013*, 59–64. Universitas Udayana.