**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

# KORELASI TIME TO LIVE TERHADAP *QUERY* TIDAK NORMAL PADA DNS MENGGUNAKAN *BINARY LOGISTIC REGRESSION*

**Aminudin[1*], Eko Budi Cahyono[1]**
[1]Program Studi Informatika, Universitas Muhammadiyah Malang
*email*: [*]aminudin2008@umm.ac.id

**Abstract:** DNS plays a vital role in the operation of services on the internet. Almost all services on the internet are under DNS control, such as email, ftp, web apps, etc. So, it is not surprising that various malicious activities involve DNS services such as financial fraud, phishing, malware and malicious activity etc. Fortunately, in DNS there is a record with the name time to live which can be used to detect a query or the address accessed from the user is a normal query or an abnormal query. Therefore, the purpose of this study is to determine the correlation value between time to live and abnormal queries on passive DNS data using the Binary Logistic Regression model. The results showed that the Binary Logistic Regression method could model the correlation between TTL, elapsed and bytes which has an optimal model F1 Score of 0.9997 and also has a condition close to the ideal state by using the Precision Recall Curve (PRC) graph plot.

**Keywords:** Binary Logistic Regression; DNS Passive; Precision Recall Curve (PRC); Query Abnormal

**Abstrak:** DNS memegang peranan yang vital di dalam berjalanya service di internet. Hampir seluruh layanan di internet berada di bawah kendali DNS seperti email, ftp, app web dll. Jadi, tidak mengherankan bahwa berbagai kegiatan jahat melibatkan layanan DNS seperti financial *fraud, phising, malware* dan aktivitas *malicious* dll. Untungnya, di dalam DNS tersimpan sebuah *record* dengan nama *time to live* yang dapat digunakan untuk mendeteksi sebuah query atau alamat yang diakses dari user tersebut bersifat query normal atau query tidak normal. Oleh karena itu, tujuan penelitian ini adalah untuk mengetahui nilai korelasi antara *time to live* dengan query tidak normal pada data passive DNS dengan menggunakan model Binary Logistic Regression. Hasil penelitian menunjukkan bahwa metode Binary Logistic Regression dapat memodelkan korelasi antara TTL, elapsed dan bytes yang memiliki model optimal F1 Score sebesar 0.9997 dan juga memiliki kondisi hampir mendekati keadaan ideal dengan menggunakan plot grafik Precision Recall Curve (PRC).

**Kata kunci:** Binary Logistic Regression; *DNS Passive; Precision Recall Curve (PRC); Query Abnormal*

## PENDAHULUAN

*Domain Name System* (DNS) merupakan pemetaan nama host ke alamat fisik yang terdapat pada alamat atau url internet. DNS Server telah menjadi sumberdaya penting untuk semua service internet yang membutuhkan resolve DNS Server, sehingga apabila DNS tidak ada maka dapat dikatakan tidak akan ada kegiatan apapun di dalam dalam service internet [1]. Ada tiga komponen utama

**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

DNS yaitu pertama DNS *Resolver* yang berfungsi untuk menjawab *request* dari *client* tentang alamat yang akan dituju, kedua *Recursive DNS Server* yang bertugas untuk meneruskan pencarian DNS melalui respons balasan dari DNS Resolver, kemudian yang ketiga *Authoritative DNS Server* menangani response yang keluar dari *Recursive DNS Server*. Oleh karena itu, DNS memainkan peranan yang sangat penting di dalam pengoperasian internet, di dalam DNS ini juga menyediakan pemetaan dua arah antara nama domain dan mengidentifikasi nilai IP dalam bentuk numerik menjadi karakter. Mengingat peranan DNS yang sangat vital di dalam internet tidak mengherankan bahwa berbagai kegiatan jahat melibatkan layanan nama domain dengan satu atau cara yang lain[2][3].

Untuk menanggulangi atau mengantisipasi dari kegiatan jahat yang ada di dalam service internet yang ada di dalam DNS maka salah satu cara yang dapat dilakukan adalah dengan mengidentifikasi dan menganalisa data passive DNS dari DNS Server. *Passive* DNS merupakan basis data yang menyimpan catatan data historis DNS dari berbagai sumber. Biasanya DNS passive berisi catatan data resolusi DNS seperti lokasi, periode waktu (*timestamps*), *latency/elapsed*, *throughput/bytes* dll [4]. Kumpulan data historis tersebut memungkinkan dianalisa untuk mengetahui lebih banyak tentang alamat yang dipetakan oleh nama domain utama pada saat itu. Jika menemukan penyimpangan dari pemetaan data yang dikumpulkan dari sumber tersebut, itu dapat menunjukkan adanya *query* yang tidak normal [5][6].

Data *passive* DNS bersifat *query* normal atau *query* tidak normal sangat yang dipengaruhi oleh nilai *Time To Live* (TTL), *elapsed/latency* dan *bytes/throughput*. TTL merupakan mekanisme untuk membatasi umur data di dalam suatu jaringan, biasanya TTL diukur dalam hitungan detik lama waktu sumber daya yang tersimpan di dalam cache lokal[1][7], nilai TTL ini yang akan dijadikan acuan utama di dalam penelitian ini. *Elapsed* merupakan jeda waktu yang dibutuhkan dalam pengantaran paket data dari pengirim ke penerima. Semakin tinggi jeda waktu tersebut maka akan semakin tinggi resiko kegagalan akses. *Bytes* adalah kecepatan dalam bertukar data dengan ukuran tertentu. Dalam penelitian ini akan diukur korelasi atau hubungan antara faktor yang sudah disebutkan diatas untuk mendeteksi normal dan tidaknya sebuah *query* di dalam *passive DNS.*

Beragam kajian telah membahas Penelitian sebelumnya pernah dilakukan dengan mengidentifikasi aktivitas botnet menggunakan data passive DNS. Penelitian tersebut menggunakan proses klasifikasi untuk mengelompokan adanya *malicious* atau tidak ada dengan menggunakan satu set 36 fitur yang berbeda. Algoritma klasifikasi yang digunakan adalah KNN, *Decision Tree* dan *Random Forest* (RF)[13]. Selain itu penelitian yang lain dengan menggunakan data passive DNS yang digunakan untuk mendeteksi *query* tidak normal dengan menggunakan nilai TTL di dalam mendeteksi *malicious*. Penelitian ini menyelidiki dugaan melalui beberapa percobaan dan hasilnya menunjukkan bahwa paket jahat dapat dibedakan tidak normal dengan mengamati nilai-nilai TTL[5]. Penelitian yang lain dengan memanfaatkan analisa lalu lintas DNS pasif untuk mendeteksi keberadaan botnet di jaringan lokal dengan menggunakan algoritma *Naïve Bayes Classifier*[14].

Dari beberapa penelitian yang su-

**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

dah disebutkan bahwa penelitian yang yang menggunakan atribut *query* tidak normal di dalam passive DNS belum pernah dilakukan, hal ini mengindikasikan bahwa penelitian yang akan dilakukan masih sangat orisinil. Dan dari pengamatan penulis bahwa metode yang dipakai yaitu *Binary Logistic Regression* juga belum pernah sekalipun dipakai untuk dilakukan penelitian tentang deteksi query tidak normal pada *passive* DNS.

## METODE

### Dataset

Pada bagian ini akan dijelaskan metode untuk tahapan penelitian yang dilakukan, meliputi: 1) pengambilan dataset 2) desain eksperimental 3) analisis korelasi.

Tabel 1. Dataset Passive DNS

| No. | Nama | Atribut | Keterangan |
|---|---|---|---|
| 1. | Period | A1 | Periode pengambilan sampel query |
| 2. | Query | A2 | Proses resolusi alamat internet |
| 3. | Type | A3 | Tipe query seperti A, NS, PTR, SOA,TXT |
| 4. | Answer | A4 | Jawaban dari server DNS |
| 5. | Response | A5 | Tanggapan server DNS terhadap query |
| 6. | TTL | A6 | Mekanisme untuk membatasi umur data DNS |
| 7. | Elapsed | A7 | Waktu yang diperlukan untuk menjawab query |
| 8. | Bytes | A8 | Jumlah paket data query |
| 9. | Cluster_id | A9 | Tanggapan sistem terhadap query |
| 10. | Src | A10 | Alamat asal query |
| 11. | Dst | A11 | Alamat tujuan query |

DNS Pasif merupakan replikasi dari catatan zona DNS seperti *query*, *type*, *answer*, TTL dan atribut yang berhubungan dengannya seperti period, *response*, *elapsed*, *bytes*, *cluster_id*. Catatan zona DNS yang terdapat pada server DNS disebut dengan DNS Aktif. Dataset DNS pasif pada penelitian ini bersumber dari replikasi catatan zona DNS dari server DNS ns1.dnsanalyzer.info dan ns2.dnsanalyzer.info yang diambil secara waktu nyata.

Tabel 2. Format Dataset Passive DNS

| No | Atribut | Format | Nilai |
|---|---|---|---|
| 1. | A1 | timestamp | 2019-01-08T07:35:00 +00:00 |
| 2. | A2 | text | kasPE.DNsaNAlYz-eR.INFO |
| 3. | A3 | text | A |
| 4. | A4 | text | 159.65.132.73 |
| 5. | A5 | numeric, category | 0 (normal) |
| 6. | A6 | numeric | 1800 (second) |
| 7. | A7 | numeric | 11.22227 (micro second) |
| 8. | A8 | numeric | 233 (byte) |
| 9. | A9 | numeric, category | 0 (normal) |
| 10. | A10 | Text | 3.85.222.253 |
| 11. | A11 | Text | 159.65.132.73 |

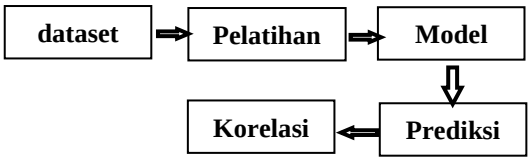
Pada penelitian ini, atribut A6, A7 dan A8 merupakan variabel bebas yang mempengaruhi atribut A9. Sedangkan atribut A9 merupakan variabel yang tengah diobservasi. Atribut A9 diobservasi langsung terhadap atribut A6 bersama-sama dengan atribut A7 dan A8. Atribut A9 (cluster_id) merupakan

**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

tanggapan sistem terhadap *query* yang sedang berjalan. Sistem akan memilah *query* menjadi dua kategori yaitu *query* normal jika cluster_id = 0 dan *query* tidak normal jika cluster_id = 1. Akurasi dari pemilahan *query* oleh sistem ini akan diuji dengan model logit sampai seberapa akurat dengan melihat tingkat korelasi antara variabel terikat sistem dengan variabel terikat model.

**Desain Eksperimental**

Pada bagian ini dijelaskan desain eksperimental model logit yang diusulkan untuk melihat korelasi antara variabel terikat sistem dengan variabel terikat model.




Gambar 1. Desain Eksperimental

Dataset disusun dalam bentuk tabel hubungan variabel bebas dengan variabel terikat. Dataset dibagi menjadi empat kelompok besar (harian, mingguan, bulanan, periode) dan total ada sebanyak enam belas kelompok uji. Pelatihan merupakan proses pembelajaran mesin dengan mengenalkan pola-pola hubungan antara variabel bebas dengan variabel terikat ke komputer sehingga komputer dapat menjawab pertanyaan hubungan antara variabel bebas dengan variabel terikat berdasarkan pola hubungan yang sudah dipelajarinya. Pola-pola hubungan tersebut adalah:

$$Log\left(\frac{PA9}{1-PA0}\right) = \beta 0 + \beta 1 * (A6) + \beta 2 * (A7) + \beta 3 * (A8) + e \quad (1)$$

Hasil yang didapat dari proses pelatihan dalam bentuk model yang mereprentasikan karakteristik hubungan antara variabel bebas dengan variabel terikat. Informasi yang terdapat pada model adalah konstanta $\beta 0$, koefisien regresi $\beta 1$, $\beta 2$, $\beta 3$ dan $e$ untuk masing-masing variabel bebas. Bentuk model berupa persamaan regresi logistik (model logit). Tahap prediksi adalah tahap untuk memetakan antara variabel bebas dengan variabel terikat dengan menggunakan model logit yang memenuhi persamaan fungsi berikut:

$$y = f(x) \quad (2)$$

Model logit merupakan fungsi binomial maka y hanya mempunyai dua nilai yaitu 0 dan 1. Dalam penelitian ini, nilai y direpresentasikan oleh atribut A9 (cluster_id). Nilai y=0 diterjemahkan untuk kondisi dimana query berada dalam kondisi normal (*TRUE*) dan nilai y=1 atau y!=0 diinterpretasikan untuk kondisi dimana *query* berada dalam kondisi tidak normal (*FALSE*). Saat pelatihan dataset dibagi menjadi dua bagian yaitu dataset latih dan dataset uji. Pada semua sampel dataset dilakukan pembagian ukuran dataset latih sebesar 50% dan dataset uji sebesar 50% secara acak. Pelatihan dilakukan terhadap dataset latih namun tidak dilakukan terhadap dataset uji untuk menjamin kemurnian proses pelatihan. Adapun beberapa keterangan terkait dengan pemrosesan data latih.

**Analisa Korelasi**

Analisis korelasi dilakukan dengan menggunakan F1 score dan *precision recall curve* (PRC) untuk mengukur tingkat korelasi antara dua variabel, dalam hal ini antara variabel terikat sistem dengan variabel terikat model.

$$F1 = 2x\frac{Precision*Recall}{Precision+Recall} \quad (3)$$

F1 Score merupakan salah satu perhitungan dalam mengukur tingkat akurasi korelasi (ketepatan dan keberhasilan) antara model dan sistem dengan

**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

mengkombinasikan *Precision* dan *Recall*. Nilai *Precision* dan *Recall* pada suatu keadaan dapat memiliki bobot yang berbeda. Ukuran yang menampilkan timbal balik antara *Precision* dan *Recall* adalah F1 score merupakan bobot *harmonic mean* dari *Precision* dan *Recall* sedangkan untuk melihat hubungan antara *Precision* dan *Recall* menggunakan grafik *Precision Recall Curve* (PRC).

## HASIL DAN PEMBAHASAN

Pada bagian ini dibahas hasil penelitian berdasarkan langkah-langkah yang telah ditetapkan pada desain eksperimental.

### Pengambilan Dataset

Tabel 3. Format Dataset Passive DNS

| Nama Sampel | Jumlah query | Mean TTL | Std dev TTL | Waktu query |
|---|---|---|---|---|
| S1 | 3.048 | 212 | 140 | 01-12-2018 s.d 02-12-2018 |
| S2 | 3.570 | 189 | 541 | 03-12-2018 s.d 04-12-2018 |
| S3 | 530 | 1.409 | 8.921 | 05-12-2018 s.d 06-12-2018 |
| S4 | 1.066 | 356 | 773 | 07-12-2018 s.d 08-12-2018 |
| S5 | 859 | 310 | 667 | 09-12-2018 s.d 10-12-2018 |
| S6 | 3.699 | 249 | 457 | 11-12-2018 s.d 12-12-2018 |
| S7 | 1.262 | 361 | 802 | 13-12-2018 s.d 14-12-2018 |
| S8 | 21.463 | 287 | 2.267 | 01-12-2018 s.d 07-12-2018 |

Dataset yang digunakan di dalam penelitian bersumber dari replikasi catatan zona DNS dari server DNS ns1.dnsanalyzer.info dan ns2.dnsanalyzer.info yang diambil secara waktu nyata. Dataset yang digunakan dimulai 01-12-2018 sampai dengan 28-02-2019 selama tiga bulan dengan data sebanyak 684.909 query. Dataset dikelompokkan menjadi empat kelompok sampel (harian, mingguan, bulanan dan periode) Atribut TTL (A6) bersama-sama dengan atribut elapsed (A7) dan bytes (A8) dipilih sebagai fitur dari variabel bebas (X) untuk mensimulasikan jaringan nyata. Atribut cluster_id (A9) ditetapkan sebagai variabel terikat (Y).

### Model

Model merupakan representasi karakteristik hubungan antara variabel bebas dengan variabel terikat. Informasi yang terdapat pada model adalah koefisien regresi w1, w2, w3 dan konstanta b yang diperoleh pada tahap pelatihan. Pada penelitian ini model digunakan untuk mencari korelasi antara nilai TTL terhadap *query* tidak normal. Karena model mempunyai kemampuan prediksi maka model dapat melakukan pengklasifikasian apakah sebuah query termasuk normal atau tidak normal. Nilai TTL berkaitan dengan nilai elapsed dan bytes yang merupakan atribut penting dalam membentuk kesatuan vektor fitur untuk merepresentasikan keadaan nyata dari jaringan.

$$z = w_1 * (TTL) + w_2 * (Elapsed) + w_3 * (Bytes) + b \quad (4)$$

Tabel 4. Model Logit Berbagai Sampel

| Sampel | w1 | w2 | w3 | b |
|---|---|---|---|---|
| S1 | -0.0219 | 0.0081 | 0.0041 | 0.2111 |
| S2 | -0.0211 | -0.0155 | 0.0086 | 0.2082 |
| S3 | -0.0572 | -0.0129 | 0.0285 | 0.1982 |
| S4 | -0.1436 | -0.0144 | 0.0304 | 0.1472 |
| S5 | -0.1265 | 0.0033 | 0.0392 | 0.1631 |
| S6 | -0.0835 | -0.0041 | 0.0333 | 0.1918 |
| S7 | -0.1345 | 0.0053 | 0.035 | 0.1553 |
| S8 | -0.0211 | 0.0021 | 0.0179 | 0.2049 |

**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

**Prediksi**

Prediksi merupakan langkah untuk memperkirakan nilai y berdasarkan model logit. Nilai y mempunyai nilai binomial yaitu 0 atau 1 yang merepresentasikan keadaan apakah suatu query normal atau tidak normal. Pada penelitian ini nilai y=0 mewakili kondisi dimana query berada dalam keadaan normal (TRUE) dan nilai y=1 atau y!=0 mewakili kondisi dimana query berada dalam keadaan tidak normal (FALSE).

Prediksi dilakukan berdasarkan model logit dengan nilai probabilitas antara 0 dan 1, padahal y adalah nilai binomial sehingga diperlukan fungsi pemeta agar semua nilai z terpetakan ke nilai binomial.

**Korelasi**

Korelasi mempunyai arti hubungan antara variabel bebas dengan variabel terikat suatu model dalam konteks regresi logistik. Hubungan yang kuat antara variabel bebas dan variabel terikat dari suatu model ditunjukkan dengan seberapa tingkat ketepatan antara model dengan sistem dan tingkat keberhasilan model memrediksi sistem. Dengan kata lain, korelasi berarti hubungan antara variabel terikat model dengan variabel terikat sistem. Dua metrik tersebut dalam statistik disebut dengan *Precision* dan *Recall*.

*Precision-Recall* menjadi instrumen ukur utama pada penelitian ini untuk menganalisis korelasi antara TTL dengan *query* tidak normal. TTL merupakan representasi dari variabel bebas dan *query* tidak normal merupakan representasi dari variabel terikat. Dalam penerapannya, TTL dikombinasikan dengan atribut *elapsed* dan bytes agar seleksi fitur mendasarkan pada lingkungan nyata suatu jaringan yang dipengaruhi oleh atribut-atribut tersebut. Sedangkan *query* tidak normal merupakan salah satu kemungkinan nilai dari variabel terikat selain *query* normal. Dalam konteks seperti ini maka nilai TTL dapat digunakan untuk memrediksi terjadinya *query* tidak normal seperti gangguan yang terjadi pada jaringan (*anomaly*) atau gangguan-gangguan lain seperti gangguan siber (*malware*). Dalam penelitian ini dipilih formula F1 Score dan *Precision Recall Curve* sebagai instrumen ukur analisis korelasi yang mendasarkan pada keseimbangan pengukuran antara *Precision* dan *Recall*.

Tabel 5. Analisa Korelasi Variable Terikat Model dan Sistem

| Sampel | $F_1$ | Precision Curve | Recall Curve |
|---|---|---|---|
| S1 | 1.0 | [1.,1.] | [1.,0.] |
| S2 | 0.995 | [0.989, 0.991, 1.] | [1., 0.999, 0.] |
| S3 | 0.986 | [0.973,1.] | [1.,0.] |
| S4 | 1.0 | [1.,1.] | [1.,0.] |
| S5 | 0.997 | [0.886, 1., 1.] | [1., 0.994, 0.] |
| S6 | 0.999 | [0.948, 1., 1.] | [1., 0.999, 0.] |
| S7 | 0.992 | [0.846, 0.988, 1.] | [1., 0.996, 0.] |
| S8 | 0.990 | [0.980,1.] | [1.,0.] |

Menurut tabel 5 rata-rata F1 harian sebesar 0,995; mingguan sebesar 0,984; bulanan sebesar 0,959; periode sebesar 0,950. Ini berarti semakin banyak dataset unjuk kerja model semakin menurun. Unjuk kerja model terbaik diperoleh pada pemodelan dataset dengan sampling harian. F1 dari sampel S1 dan S4 masing-masing sebesar 1.0 menunjukkan bahwa pelatihan model mengalami *overfitting*. Hal ini berakibat buruk terhadap model jika dilakukan pengujian dengan data yang berbeda karena akan mengurangi akurasi prediksi. Dilihat dari *Precion Recall Curve* sampel

**JURTEKSI (Jurnal Teknologi dan Sistem Informasi)** ISSN 2407-1811 (Print)
Vol. VII No. 2, April 2021, hlm. 143 - 150 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v7i2.924
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

yang mendekati kondisi model ideal adalah sampel 5, 6, 7 dengan nilai cut-off Precision berturut-turut sebesar [0.88604651,1.,1.], [0.94810811,1.,1.], [0.84627575,0.98884758,1.]; nilai cut-off Recall berturut-turut sebesar [1.,0.99475066,0.], [1.,0.99942987,0.], [1.,0.99625468,0.].

Berdasarkan hasil analisis korelasi variabel terikat model dan sistem dapat disimpulkan bahwa model optimal dicapai oleh model yang dihasilkan dari pelatihan sampel dataset ke 6 (S6). Ini berarti, dari model dapat diketahui adanya korelasi kuat antara TTL dengan query tidak normal, dan atau sebaliknya terhadap query normal. Korelasi tersebut dapat dirumuskan:

$$y = -0.083 * \text{ttl} + -0.004 * \text{elapsed} + 0.033 * \text{bytes} + 0.191.$$

## SIMPULAN

*Precision-Recall* menjadi instrumen ukur utama pada penelitian ini untuk menganalisis korelasi antara TTL dengan query tidak normal. *Query* tidak normal merupakan kemungkinan nilai dari variabel terikat selain query normal. Dalam konteks seperti ini maka nilai TTL dapat digunakan untuk memrediksi terjadinya *query* tidak normal seperti gangguan yang terjadi pada jaringan (*anomaly*) atau gangguan-gangguan lain seperti gangguan siber (*malware*). Regresi logistik dapat memodelkan korelasi antara ttl, elapsed dan bytes dengan query normal atau query tidak normal yang memenuhi persaman regresi $y = -0.083 * \text{ttl} + -0.004 * \text{elapsed} + 0.033 * \text{bytes} + 0.191$ yang artinya Semakin banyak dataset unjuk kerja model semakin menurun. Model optimal diperoleh dengan $F_1$ Score sebesar 0.9997 dan kondisi hampir mendekati keadaan ideal terlihat pada plot grafik *Precision Recall Curve* (PRC).

## DAFTAR PUSTAKA

[1] I. Van Zyl and B. Irwin, "A review of current DNS TTL practices," no. September 2015, 2018.

[2] C. N. Cs, D. N. S. Overview, T. Dns, T. Rfc, and A. Dns, "DNS Packet Structure," no. September, 2009.

[3] S. Torabi, A. Boukhtouta, C. Assi, and M. Debbabi, "Detecting internet abuse by analyzing passive DNS traffic: A survey of implemented systems," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 4, pp. 3389–3415, 2018, doi: 10.1109/COMST.2018.2849614.

[4] LastLine, "Using Passive DNS Analysis to Automatically Detect Malicious Domains."

[5] A. Alenazi, "HTTP Botnet Detection using Passive DNS Analysis and Application Profilin," Vancouver Island University, 2015.

[6] A. M. Kara, H. Binsalleeh, M. Mannan, A. Youssef, and M. Debbabi, "Detection of malicious payload distribution channels in DNS," *2014 IEEE Int. Conf. Commun. ICC 2014*, pp. 853–858, 2014, doi: 10.1109/ICC.2014.6883426.

[7] D. Wielogorska, Monika; O'Brien, "DNS Analysis for Botnet Detection," vol. 550, no. Spring, pp. 1–8, 2014.

[8] S. Marchal *et al.*, "DNSSM: A Large Scale Passive DNS Security Monitoring Framework," pp. 988–993, 2012, doi: 10.1145/1064212.1064271.

[9] R. Yamada and S. Goto, "Using abnormal TTL values to detect malicious IP packets," pp. 3–4, 2012.

[10] X. Li, J. Wang, and X. Zhang, "Botnet detection technology based on DNS," *Futur. Internet*, vol. 9, no. 4, pp. 1–12, 2017, doi: 10.3390/fi9040055.

[11] W. Putera, *Using Logistic Regression Method for Analysis Voting Behaviour in Political Science*. .

[12] N. D. Sinaga, "Model Regresi Logistik Biner untuk Menentukan Faktor yang Berpengaruh Terhadap Anak Putus Sekolah di Sulawesi Tengah," vol. 13, no. 1, pp. 24–37, 2016.

[13] Y. Wijaya,Arianto;Darsyah, "Binary Logistic Regression (BLR) untuk Mengetahui Pengaruh Tingkat Pendidikan dan Jenis Kelamin Terhadap Status Bekerja di Kota Surabaya," no. 1, pp. 3–10, 2005, doi: 10.4135/9781412995627.

[14] R. Hendayana, "Penerapan Metode Regresi Logistik dalam Menganalisis Adopsi Teknologi Pertanian," *Inform. Pertanan*, vol. 22, no. 2, pp. 1–9, 2012.

[15] S. Hosmer, David; Lemeshow, *Applied Logistic Regression*, Second. 2000.

[16] P. da Pedro Marques Luz, "Botnet Detection Using Passive DNS," 2014.