

ANALISIS PERBANDINGAN DETECTION TRAFFIC ANOMALY DENGAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)

Imam Riadi¹, Rusydi Umar², Fadhilah Dhinur Aini³

¹imam.riadi@is.uad.ac.id, ²rusydi_umar@rocketmail.com, ³fadhilah1708048025@webmail.uad.ac.id
Program Studi Sistem Informasi¹, Program Studi Teknik Informatika^{2,3}
Universitas Ahmad Dahlan

Abstrak

Intrusion Detection System (IDS) merupakan sebuah perangkat lunak atau perangkat keras yang dapat digunakan untuk mendeteksi adanya aktivitas yang tidak wajar dalam jaringan. Situasi sering muncul dari berbagai akses jaringan berupa informasi atau data yang dapat menimbulkan masalah. Deteksi merupakan sistem untuk mendeteksi aktivitas yang bersifat mengganggu akses data dalam sebuah informasi. IDS memiliki dua metode dalam melakukan pendeteksian yaitu Rule Based (Signature Based) dan Behavior-Based. Traffic Anomaly dapat mendeteksi peningkatan jumlah akses pengguna dan sewaktu – waktu akan terjadi sebuah serangan dari pihak lain terhadap jaringan tersebut. Penelitian ini menggunakan 2 Metode algoritma yaitu Naïve Bayes dan Support Vector Machine (SVM). Hasil Naïve Bayes melalui sampel data grafik Distributions dan Radviz memiliki nilai probabilitas 0.1 dan nilai probabilitas paling tinggi yaitu 0.8. Untuk Support Vector Machine (SVM) menghasilkan grafik yang memiliki lebih besar nilai akurasinya.

Kata kunci: Klasifikasi Naive Bayes, Support Vector Machine (SVM), *Intrusion Detection System* (IDS), Traffic Anomaly

Abstract

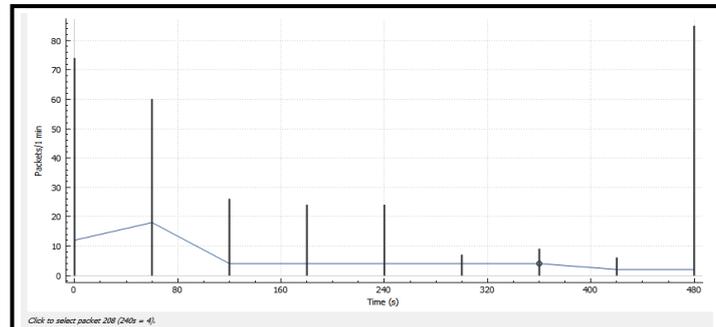
Intrusion Detection System (IDS) is a software or hardware that can be used to detect any abnormal activity in the network. Situations often arise from various network access in the form of information or data that can cause problems. Detection is a system for detecting activities that are disturbing data access in information. IDS has two methods of doing detection, namely Rule Based (Signature Based) and Behavior-Based. Anomaly traffic can detect an increase in the number of user access and at any time there will be an attack from another party on the network. This study uses 2 algorithm methods are Naïve Bayes and Support Vector Machine (SVM). Naïve Bayes results through the Distributions and Radviz graph data samples have a probability value of 0.1 and the highest probability value is 0.8. Support Vector Machine (SVM) produces a graph that has greater accuracy.

Keywords: Classification Naive Bayes, Support Vector Machine (SVM), *Intrusion Detection System* (IDS), Traffic Anomaly

1. Pendahuluan

Intrusion Detection System (IDS) merupakan sistem perangkat lunak atau perangkat keras yang dapat digunakan untuk mendeteksi adanya aktivitas yang mencurigakan dalam sistem jaringan komputer [1]. Dalam permasalahan IDS ini telah didekati oleh beberapa algoritma seperti *Naïve Bayes*, support vector machine (SVM) dan algoritma lainnya. Naive Bayes adalah sebuah metode atau algoritma klasifikasi sederhana yang mampu berkontribusi pada keputusan akhir dan pada setiap atributnya memiliki sifat independent. *Naïve Bayes* merupakan salah satu metode di dalam data mining untuk mengklasifikasikan data. Cara kerja dari metode Naïve Bayes menggunakan parameter yang telah ada. Konsep dasar Naïve Bayes adalah Teorema Bayes. Teorema yang digunakan dalam statistika untuk menghitung suatu peluang, Bayes Optimal Classifier menghitung peluang dari satu kelas dari masing – masing kelompok atribut yang ada dan menentukan kelas mana yang paling optimal. Proses pengelompokan atau klasifikasi dibagi menjadi dua fase yaitu *learning/training* dan *testing/classify* Pada fase *learning*, sebagian data yang telah diketahui kelas, datanya diumpangkan untuk membentuk model perkiraan. Kemudian pada fase *testing*, model yang sudah terbentuk diuji dengan sebagian data [2]. Traffic Anomaly merupakan suatu keadaan yang tidak stabil terjadi di lalu lintas jaringan sehingga rentannya sebuah jaringan untuk di serang. Anomali trafik tersebut dapat melumpuhkan jaringan dari sisi file yang target dari penyusup [3].

Traffic Anomaly dapat mendeteksi peningkatan jumlah akses pengguna dalam sewaktu waktu akan terjadi sebuah serangan dari pihak lain terhadap jaringan dapat dilihat pada gambar 1.



Gambar 1. Deteksi Traffic Anomaly

Gambar 1 diatas merupakan data grafik menunjukkan pengiriman packet dan penerimaan packet menggunakan akses jaringan wireless.

2. Metode

Adapun beberapa penelitian yang terkait yang pernah dilakukan oleh Penelitian [4] membahas tentang “Perbandingan IDS Snort dan IDS Suricata dalam mendeteksi serangan TCP SYN Flood”, mendapatkan hasil yang positif. Dengan terbukti kedua hasil IDS dalam mendeteksi aktivitas tanpa harus mengalami kendala. Pada penelitian tersebut disimpulkan bahwa metode Intrusion Detection System (IDS) Snort lebih unggul dibandingkan dengan Intrusion Detection System (IDS) Suricata serta mempunyai nilai akurasi lebih tinggi dibandingkan Intrusion Detection System (IDS) Suricata lebih rendah

Penelitian [5] membahas tentang “Implementasi Metode Support Vector Machine untuk melakukan Klasifikasi Kemacetan lalu lintas pada twitter”, dalam penelitian ini melakukan scenario yang telah dilakukan dapat dilihat bahwa tinggi rendahnya akurasi dipengaruhi oleh jumlah dataset yang digunakan. Hasil pengujian tersebut dapat dirata – rata akurasi tertinggi yaitu 98.67% pada data jumlah dataset 934.

Penelitian [6] membahas tentang “Sistem Deteksi Intrusi dengan Snort”, sistem yang menghasilkan sebuah log yang tersimpan dalam *database* dan alert yang dihasilkan dapat ditampilkan dan dianalisis dalam tampilan web.

Penelitian [7] membahas tentang “Analisa dan Perbandingan Bukti Forensik Aplikasi Media Sosial Facebook dan Twitter Pada Smartphone Android”, penelitian ini dilakukan untuk menemukan dan membanding bukti-bukti forensic dengan menjalankan 11 skenario diantaranya adalah pengembalian file yang dihapus, pencarian bukti forensic berupa nama akun, lokasi, nomor telepon, tanggal lahir, photo profile, cover photo, posting.

Penelitian [8] membahas tentang “Analisis Live Forensik Untuk Perbandingan Keamanan Email Pada Sistem Operasi Proprietary”, menghasilkan eksperimen yang dilakukan dengan menggunakan Personal Computer Sistem Operasi Windows 10 64bit, browser Mozilla Firefox 49.0.1. Email ini merupakan akun yang terintegrasi dengan akun sosial media lain untuk itu harus terjaga keamanannya. Metode live forensics merupakan suatu teknik untuk menemukan barang bukti pada data volatile termasuk username dan password. Jasa penyedia email terus berkembang dengan menambahkan berbagai fitur demi kenyamanan pengguna termasuk fitur keamanan.

Penelitian [9] membahas tentang “Analisis Statistik Log Jaringan Untuk Deteksi Serangan DDOS berbasis Neural Network”, mendeteksi serangan DDoS dengan metode neural network dengan fungsi *Fixed Moving Average Window (FMAW)* menghasilkan presentase rata-rata pengenalan terhadap 3 kondisi jaringan yaitu normal, slow DDoS, Dan DDoS sebesar 90,52%.

Naive Bayes merupakan salah satu metoda *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naive bayes proses klasifikasi statistik yang bisa digunakan dalam melakukan prediksi suatu probabilitas pada keanggotaan sebuah *class* [10].

Support Vector Machine merupakan sebuah metode yang membandingkan suatu seleksi parameter standart nilai diskrit yang disebut kandidat set. Untuk mengklasifikasikan akurasi Support Vector Machine (SVM) diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992.

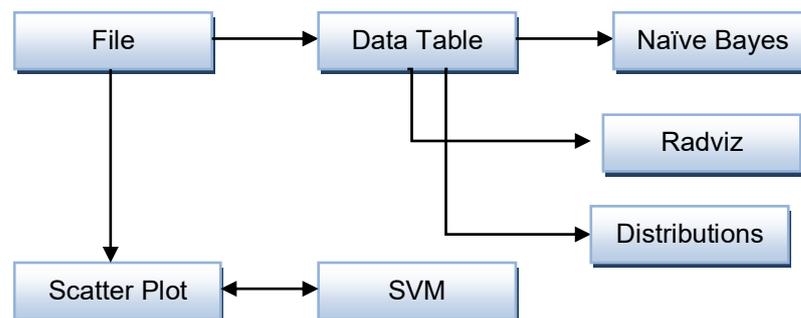
Metode Support Vector Machine memiliki 5 komponen yang berfungsi di antaranya:

- SVM Linear
- SVM Polynomial
- Kernel RBF (*Radio Basis Function*)
- Kernel MLP (*Multi Layer Perceptrom*)
- Tangent Hyperbolic (*sigmoid*)

Metode Support Vector Machine *Sigmoid* merupakan sebuah parameter yang dibuat untuk memudahkan dalam pemecahan klasifikasi data yang dihubungkan ke scatter plot untuk mendapatkan hasil titik deteksi yang diakses oleh para pengguna dari IP address [11]

Intrusion Detection System merupakan sebuah perangkat keras maupun perangkat lunak yang mampu melakukan suatu deteksi pada suatu aktifitas yang mencurigakan yang terjadi pada jaringan komputer. IDS mempunyai beberapa kategori yaitu *Network-based Intrusion Detection System (NIDS)* dan *Host-based Intrusion Detection System (HIDS)*. [12]

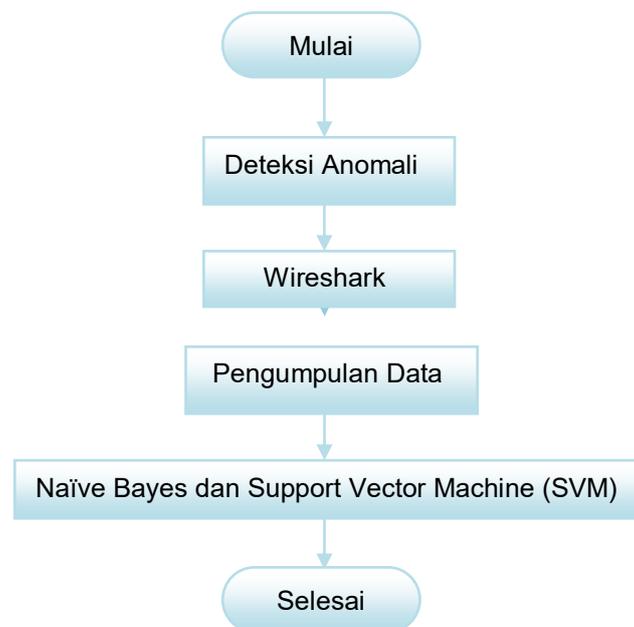
Data *traffic* yang diproses melalui klasifikasi Naïve Bayes dan SVM yang selanjutnya dilakukan pengujian hingga menghasilkan nilai dari Data table, Distributions, Radviz, Scatter Plot dapat dilihat gambar 2.



Gambar 2. Alur Pengujian

Merupakan alur pengujian yang memproses semua data yang sudah di inputkan kedalam file dengan menghubungkan ke data table dan scatter plot, lalu data table dihubungkan ke Naïve bayes, Radviz, Distribusi. Terakhir Support Vector Machine (SVM) dihubungkan ke Scatter Plot untuk menghasilkan sebuah grafik.

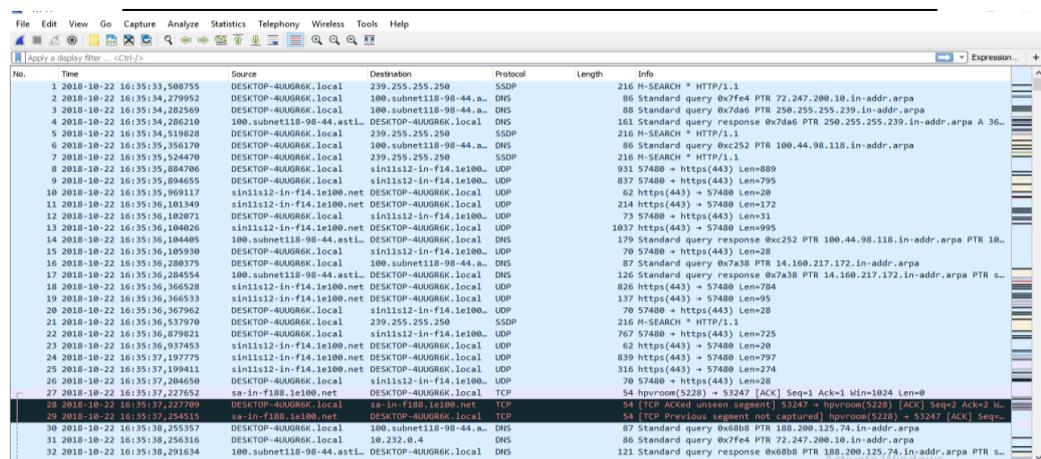
Tahap-tahap yang dilakukan dalam penelitian ini dalam bentuk diagram alir gambar 3.



Gambar 3. Diagram Alir Penelitian

Penjelasan diagram alir yang ditunjukkan pada gambar 3 sebagai berikut:

- a) Deteksi Anomaly Traffic
Anomaly detections merupakan suatu monitoring untuk memantau pergerakan yang terjadi pada sistem jaringan. Jika terjadi penyerangan terhadap sistem maka anomaly traffic akan mendeteksi jumlah peningkatan pada jaringan tersebut.
- b) Pencarian data dilakukan melalui Wireshark yang terhubung dengan jaringan internet untuk mengetahui adanya akses *Traffic Anomaly* yang mencurigakan seperti paket yang berwarna hitam gambar 4.



Gambar 4. Pencarian Data

- c) Data merupakan proses suatu pengumpulan dari hasil capture melalui wireshark yang terhubung dengan akses jaringan internet lalu dikonfigurasi dalam bentuk .csv, sehingga dapat diolah lebih lanjut menggunakan kedua metode tersebut gambar 5.

Gambar 5. Pengumpulan Data

- d) Metodologi penelitian perbandingan Naïve Bayes dan SVM. Naïve Bayes Classifier disebut sebagai multinomial naïve bayes merupakan model penyederhanaan dari algoritma bayes yang cocok dalam pengklasifikasikan text atau dokumen [13].

Persamaan rumus Navie Bayes sebagai berikut:

$$P(A|B) = (P(B|A) * P(A))/P(B) \dots\dots\dots(1)$$

Peluang kejadian A bersyarat B ditentukan dari peluang B saat A, peluang A dan Peluang B berubah menjadi.

$$P(A|B) = (P(D|Ci) * P(Ci)) / P(D) \dots\dots\dots(2)$$

$$V_{MAP} = \arg \max PV_j | a_1, a_2, \dots, a_n \dots\dots\dots (3)$$

Keterangan:

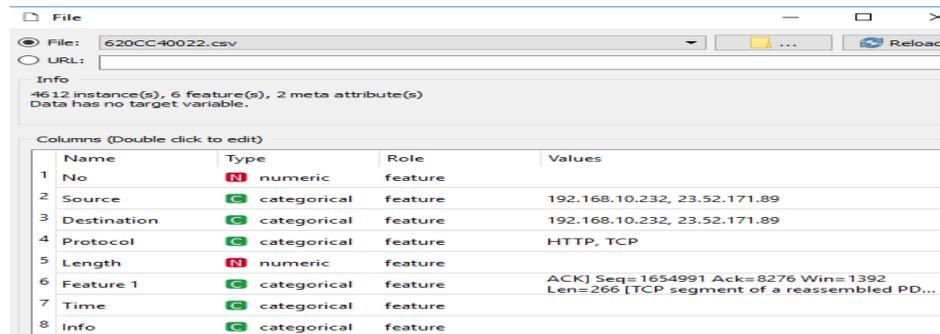
P(V_j) : Probabilitas setiap dokumen terhadap sekumpulan dokumen.

P(W_k|V_j) : Probabilitas kemunculan kata W_k pada suatu dokumen dengan kategori class V_j

- e) Metode *Support Vector Machine* (SVM) merupakan suatu parameter atau atribut untuk mengklasifikasikan data kedalam set pengujian. Pengujian dilakukan dalam satu set berisi nilai target serta beberapa fitur lainnya. Tujuan SVM adalah untuk menghasilkan model yang dihubungkan dengan *Scatter Plot* yang menghasilkan sebuah grafik perbandingan data [14]. *Scatter Plot* merupakan sebuah grafik yang bisa digunakan untuk melihat suatu pola hubungan antara 2 variabel. Untuk bisa menggunakan scatter plot, skala data harus digunakan dalam skala interval dan rasio [15].

4. Hasil dan Pembahasan

Analisis perbandingan menggunakan Naïve Bayes dan Support Vector Machine diperoleh bahwa hasil nilai akurasi probabilitas tertinggi dari data akses TCP dengan proses simulasi pengujian data yang di akses melalui *traffic anomaly* yang diakses melalui tools *wireshark* dan menghasilkan format data dalam bentuk *csv*.



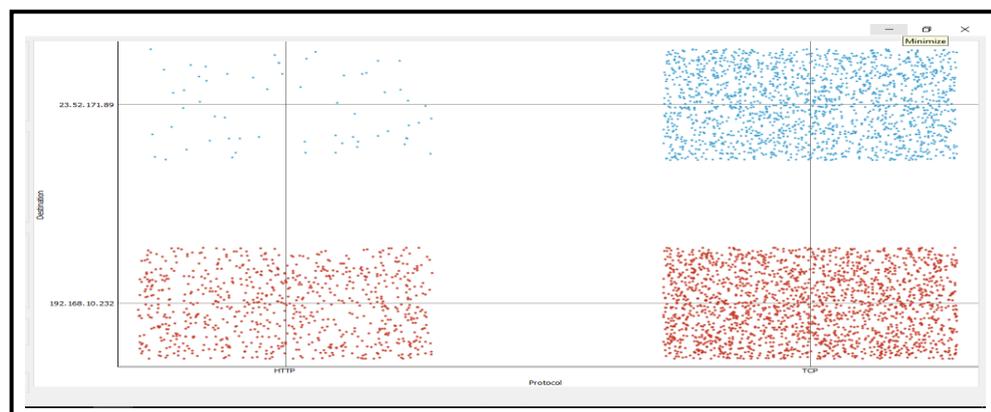
Gambar 6. Inputan Dokumen File

Gambar 6 merupakan hasil inputan dokumen / file yang akan di proses dengan data table yang akan di hasilkan pada Gambar 7 sebagai berikut :

	No	Source	Destination	Protocol	Length
1	3	23.52.171.89	192.168.10.232	HTTP	1514
2	6	23.52.171.89	192.168.10.232	HTTP	1514
3	10	192.168.10.232	23.52.171.89	TCP	54
4	13	23.52.171.89	192.168.10.232	HTTP	1514
5	20	192.168.10.232	23.52.171.89	TCP	66
6	28	23.52.171.89	192.168.10.232	TCP	1514
7	31	192.168.10.232	23.52.171.89	TCP	54
8	39	23.52.171.89	192.168.10.232	HTTP	320
9	42	192.168.10.232	23.52.171.89	TCP	54
10	723	192.168.10.232	23.52.171.89	HTTP	385
11	1272	23.52.171.89	192.168.10.232	TCP	1514
12	1273	23.52.171.89	192.168.10.232	TCP	1514
13	1274	192.168.10.232	23.52.171.89	TCP	54
14	1286	23.52.171.89	192.168.10.232	TCP	1514
15	1287	23.52.171.89	192.168.10.232	TCP	1514
16	1288	192.168.10.232	23.52.171.89	TCP	54
17	1309	23.52.171.89	192.168.10.232	TCP	1514
18	1310	23.52.171.89	192.168.10.232	TCP	1514
19	1318	192.168.10.232	23.52.171.89	TCP	54
20	1342	23.52.171.89	192.168.10.232	TCP	1514
21	1343	23.52.171.89	192.168.10.232	TCP	1514
22	1349	192.168.10.232	23.52.171.89	TCP	66

Gambar 7. Data traffic

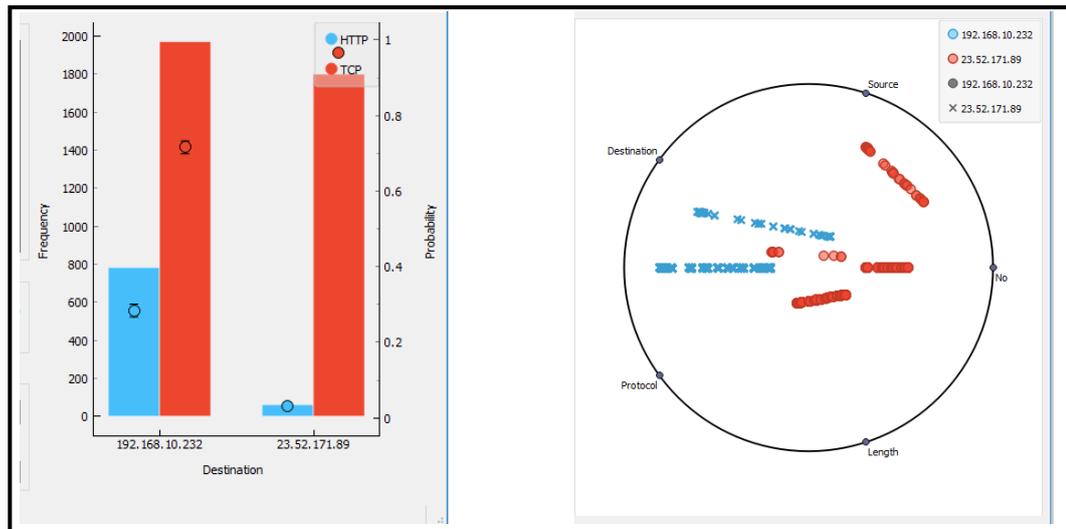
Untuk proses selanjutnya akan dihasilkan melalui Scatter Plot, Ridviz dan Distributions yang berupa grafik pada gambar 8 menghasilkan grafik Scatter Plot yang telah diproses melalui SVM sebagai berikut :



Gambar 8. Hasil Grafik Scatter Plot

Gambar 8 menghasilkan grafik scatter plot ini terbagi 4 grafik yang digunakan untuk melihat suatu pola dengan skala data yang digunakan skala interval, rasio, dan penyebaran data. Untuk hasil pada warna biru pertama scatter plot menganalisis penyebaran data dari hasil SVM menunjukkan pola grafik menyebar lebih sedikit, untuk warna biru kedua penyebaran data menunjukkan hasil grafik banyaknya pengumpulan data menggunakan TCP. Sedangkan grafik untuk warna merah pertama penyebaran data menggunakan HTTP mempunyai data yang banyak menyebar ke suatu area dan warna merah kedua mempunyai pola grafik kumpulan datanya lebih banyak terkumpul dalam satu area.

Proses selanjutnya hasil perbandingan Naïve bayes melalui Grafik Distributions dan Radviz sebagai berikut:



Gambar 9 Grafik Distributions

Gambar 10. Grafik Radviz

Hasil Naïve Bayes melalui grafik *Distributions* dan *Radviz* gambar 10. Gambar 9 menghasilkan grafik warna biru *HTTP* presentase dengan nilai akurasi 800 frekuensi dengan probabilitas 0.4 dan warna merah *TCP* kedua nilai akurasinya probabilitas 0.1. Sedangkan warna merah *HTTP* memiliki nilai akurasi 1900 frekuensi dan nilai probabilitasnya *TCP* melebihi dari 0.8. Hasil dari grafik radviz dengan IP 192.168.10.232 berwarna biru dengan penyebaran data lebih banyak dibandingkan dengan IP 23.52.171.89 penyebaran datanya lebih sedikit.

5. Kesimpulan dan Saran

Berdasarkan analisis perbandingan menggunakan Naïve Bayes dan Support Vector Machine diperoleh bahwa nilai akurasi yang di hasilkan oleh Naïve Bayes melalui sampel data grafik *Distributions* dan *Radviz* memiliki nilai probabilitas 0.1 dan nilai probabilitas paling tinggi yaitu 0.8 sedangkan hasil Support Vector Machine (SVM) memiliki nilai akurasi yang paling banyak penyebaran menggunakan Scatter Plot. Nilai yang maksimal dalam analisis perbandingan menggunakan metode Naïve Bayes.

Saran mengenai proses dari analisis perbandingan tersebut dapat dilakukan dengan proses perbandingan dengan metode - metode lain dan menggunakan aplikasi, tools yang memberi para pengguna untuk mengembangkan hal tersebut jauh lebih baik dari sebelumnya.

Daftar Pustaka

- [1] M. Jannah, Hustinawati, and R. Wildani, "Implementasi Intrusion System (Ids) Snort Pada Laboratorium Jaringan Komputer," *UG J.*, vol. 6 No 5, pp. 1–4, 2012.
- [2] M. Sudarma and D. P. Hostadi, "Komunikasi Pada Network Traffic Menggunakan Naïve Bayes Sebagai," *Icsgteis*, no. November, pp. 59–64, 2013.
- [3] Y. Purwanto and F. Y. Suratman, "Perancangan Dan Analisis Deteksi Anomali Berbasis Clustering Menggunakan Algoritma Modified K-Means Dengan Timestamp Initialization Pada Sliding Window Design And Analysis Of Anomaly Detection Based Clustering Using Modified K-Means Algorithm With Timesta."
- [4] E. Risyad, M. Data, and E. S. Pramukantoro, "Perbandingan Performa Intrusion Detection System (IDS) Snort Dan Suricata Dalam Mendeteksi Serangan TCP SYN Flood," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, pp. 2615–2624, 2018.
- [5] E. Susilowati, M. K. Sabariah, and A. A. Gozali, "Implementasi Metode Support Vector Machine Untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter Implementation Support Vector Machine Method for Traffic Jam Classification on Twitter," *E-Proceeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.

- [6] A. P. Wicaksono, J. Raya, D. Po, and B. Purwokerto, "Sistem Deteksi Intrusi dengan Snort (Intrusion Detection System with Snort)," vol. III, pp. 31–34, 2014.
- [7] W. A. Mukti, S. U. Masruroh, D. Khairani, and B. Forensik, "Analisa Dan Perbandingan Bukti Forensik Aplikasi Media Sosial Facebook Dan Twitter Pada Smartphone Android," vol. 10, no. 1, pp. 73–84, 2017.
- [8] M. N. Faiz, R. Umar, and A. Yudhana, "Analisis Live Forensics Untuk Perbandingan Keamanan Email Pada Sistem Operasi Proprietary," *J. Ilm. Ilk.*, vol. 8, no. 3, pp. 242–247, 2016.
- [9] A. W. Muhammad, I. Riadi, and Sunardi, "Analisis Statistik Log Jaringan Untuk Deteksi Serangan Ddos Berbasis Neural Network," *J. Ilm. Ilk.*, vol. 8, no. Desember, pp. 220–225, 2016.
- [10] I. N. T. Wirawan and I. Eksistyanto, "Penerapan Naive Bayes Pada Intrusion Detection System Dengan Diskritisasi Variabel," *J. Ilm. Teknol. Inf.*, vol. 13, pp. 182–189, 2015.
- [11] N. T. Thomopoulos, *Statistical Distributions*. 2017.
- [12] J. Gondohanindijo, "Sistem Untuk Mendeteksi Adanya Penyusup (IDS : Intrusion Detection System)," *Semarang*, vol. 2, pp. 46–54, 2011.
- [13] Y. S. Nugroho, "Data Mining Menggunakan Algoritma Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant.* 2013, pp. 1–11, 2009.
- [14] C. Yang, G. N. Odvody, C. J. Fernandez, J. A. Landivar, R. R. Minzenmayer, and R. L. Nichols, "Evaluating unsupervised and supervised image classification methods for mapping cotton root rot," *Precis. Agric.*, vol. 16, no. 2, pp. 201–215, 2015.
- [15] S. P. H. Pb, "Scatterplots and Correlation," *Growth (Lakeland)*, 2003.