Jurnal Kreativitas Mahasiswa Informatika
Volume 1 Nomor 2 Tahun 2020 Page 99 - 103.
p-ISSN: 2797-6327 e-ISSN:

ANALISIS PERBANDINGAN DETECTION TRAFFIC ANOMALY DENGAN METODE NAIVE BAYES DAN DBSCAN

Rino Bahtiar[1], Moch. Dwi Sakti T[2], Aris Setiawan[3], Perani Rosyani[4]
[1,2,3,4]Universitas Pamulang; Jl. Raya Puspitek No. 46 buaran, serpong, Kota Tangerang Selatan. Provinsi Banten 15310. (021) 741-2566 atau 7470 9855
[1,2,3,4]Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pamulang

e-mail: [1]rinbahtiar357@gmail.com, [2]muhammaddwi79@gmail.com , [3]aris.seti990@gmail.com, [4]dosen00837@unpam.ac.id

Abstrak

Intrusion Detection System (IDS) adalah sebuah perangkat lunak atau 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 adalah sebuah 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. Pada penelitian ini menggunakan 2 Metode algoritma yaitu Naïve Bayes dan DBSCAN Menggunakan *streaming traffic*. Hasil Naïve Bayes melalui sampel data grafik *Distributions* dan *Radviz* memiliki nilai probabilitas 0.1 dan nilai probabilitas paling tinggi yaitu 0.8. dan DBSCAN memiliki performansi yang baik dalam mendeteksi anomali trafik. Hal ini dapat ditunjukkan dengan pengujian yang dilakukan terhadap Akurasi dari hasil klaster, dimana nilai rata rata akurasinya adalah 98.45 % serta memakan waktu kurang lebih 600 detik atau sekitar 10 menit dalam sekali proses 30.000 data.

Kata kunci: Klasifikasi Naive Bayes, DBSCAN, Intrusion Detection System (IDS), Traffic Anomaly

I. PENDAHULUAN

Pertumbuhan pesat pada jaringan teknologi yang dialami seluruh duniamemberikan dampak positif dalam pengolahan informasi dan data. Pada umumnya jaringan internet juga memiliki dampak *negative* yang belum banyak diketahui oleh para pengguna, ini diakibatkan oleh pihak-pihak yang tidak bertanggung jawab dengan mengambil keuntungan untuk kepentingan mereka sendiri. *Intrusion-Detection System* (IDS) adalah sebuah cara yang dapat digunakan untuk mendeteksi aktivitas yang mencurigakan dalam sebuah sistem atau jaringan[1]. *Intrusion detection system* (IDS) berfungsi untuk mengidentifikasi lalu lintas paket-paket data yang ditransmisikan melalui jaringan computer, selanjutnya menentukan paket-paket data tersebut aman, mencurigakan atau merupakan sebuah serangan[2].

Algorima Naive Bayes adalah algoritma yang sederhana dan tiap-tiap atribut bersifat bebas, yang memungkinkan tiap atribut dapat berkontribusi terhadap hasil akhir. salah satu metode algoritma yang termasuk pada teknik klasifikasi adalah algoritma Naïve Bayes[3]. Konsep dasar Naïve Bayes adalah Teorema Bayes. Teorema yang digunakan dalam statiska 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[4].

DBSCAN adalah metode yang menggunakan konsep titik pusat (*core point*), titik batas (*border point*), dan noise. Titik yang memiliki sejumlah titik tetangga dan memenuhi jumlah titik minimum, serta berada dalam jarak tertentu disebut sebagai titik pusat, sedangkan titik batas memiliki jumlah titik tetangga namun tidak memenuhi jumlah titik minimum. Titik batas tersebut biasanya merupakan titik di dalam ketetanggaan dari

Jurnal Kreativitas Mahasiswa Informatika
Volume 1 Nomor 2 Tahun 2020 Page 99 - 103.
p-ISSN: 2797-6327 e-ISSN:

titik pusat. Kriteria suatu titik dikatakan sebagai noise yaitu pada saat titik tersebut tidak termasuk titik pusat maupun titik batas, selain itu titik tersebut tidak memenuhi konsep directly density-reachable dari suatu titik pusat (Ester et al. 1996)[5].

Clustering adalah salah satu alat data mining, pengelompokan penggunaan untuk membagi data ke dalam cluster yang bermakna atau berguna. Sebagian besar algoritma umum gagal untuk menghasilkan hasil yang berarti karena sparsity yang melekat pada benda. Dengan dimensi data yang tinggi, dan distribusi bola super, cenderung gagal dalam data terorganisir property karakteristik pengelompokan data[5].

Deteksi serangan pada saat ini lebih banyak menggunakan data non real time, artinya data masih bersifat offline dengan cara di *damped*. Sehingga ketika diaplikasikan ke data yang bersifat realtime akan menimbulkan perbedaan yang signifikan. Dalam pengembangannya sistem deteksi anomali banyak pendekatan untuk mengetahui pola trafik normal sebagai acuan deteksi anomali trafik. Penelitian ini menggunakan algoritma Naïve Bayes dan DBSCAN dalam mengelompokkan trafik apakah tergolong pada trafik normal ataukah trafik anomali[6].

II. METODE PELAKSANAAN

Pada penelitian ini dilakukan percobaan untuk mendeteksi *anomaly* pada jaringan menggunakan Algoritma *Naïve Bayes* dan *DBSCAN*. Pada penelitian ini penulis menyusun desain penelitian berupa tahapan-tahapan penelitian agar penelitian dapat dilakukan secara sistematis. dan terakhir melakukan perbandingan nilai akurasi Antara Algoritma Naïve Bayes dan *DBSCAN*.

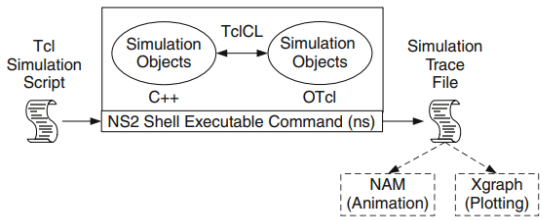
1. Deteksi Anomali Trafik

Anomali trafik adalah suatu keadaan yang menyebabkan terjadinya abnormalitas pada lalu lintas jaringan. Penyebab dari anomali ini bias dari banyaknya faktor seperti pengguna internet atau serangan pada suatu jaringan. Kondisi ini dapat mengakibatkan penurunan performansi jaringan sehingga rentannya sebuah jaringan untuk diserang. Dampak dari anomali trafik ini dapat melumpuhkan jaringan, dapat juga disisipi file yang membahayakan target dari penyerang. Maka dari itu, diperlukan pendeteksi lalu lintas jaringan untuk mencegah ataupun mengatasi anomali didalam lalu lintas jaringan.

Pada Deteksi anomali trafik terdapat dua istilah pendekatan yaitu *Intrusion Detection System* (IDS) dan *Intrusion Prevention System* (IPS). IDS dan IPS memiliki sistem yang bekerja dalam mengawasi keadaan jaringan serta memberikan peringatan kepada administrator apabila mendeteksi serangan datang ataupun anomaly. Pada IDS/IPS menggunakan metode *traffic anomaly based* dan *intrusion signature*. Metode *intrusion signature* dapat mendeteksi serangan berdasarkan *database* yang dimiliki, sehingga kelemahannya jika sebuah serangan tidak ada di database, maka akan dianggap sebagai trafik normal. Pada *traffic anomaly based*, proses pendeteksian tidak membutuhkan serangan pada *database*, sehingga proses pendeteksian tidak bergantung pada *database.*

2. Network Simulator – 2 (NS2)

Network Simulator – 2 (NS2) terdiri dari dua bahasa utama: CCC dan Perintah Alat Berorientasi Objek Bahasa (OTcl). Sementara CCC mendefinisikan mekanisme internal (yaitu, backend) simulasi, OTcl menyiapkan simulasi dengan merakit dan mengonfigurasi objek serta penjadwalan acara diskrit (yaitu, sebuah frontend). CCC dan OTcl dihubungkan bersama menggunakan TclCL. Dipetakan ke objek CCC, variable dalam domain OTcl kadang-kadang disebut sebagai pegangan. Secara konseptual, pegangan hanya berupa string (mis., "_o10") dalam domain OTcl dan tidak mengandung apa pun Kegunaan. Sebaliknya, fungsionalitas (misalnya, menerima paket) didefinisikan dalam objek CCC yang dipetakan (mis., Konektor kelas). Dalam domain OTcl, sebuah pegangan bertindak sebagai antarmuka yang berinteraksi dengan pengguna dan objek OTcl lainnya. Mungkin mendefinisikan prosedur dan variabelnya sendiri untuk memfasilitasi interaksi.



Gambar 2. 1 Arsitektur Dasar NS

3. *Preprocessing*

Preprocessing diperlukan untuk mentransformasi raw data menjadi data yang mudah diinterpretasikan sebagai inputan algoritma sistem deteksi anomali untuk dianalisis. *Preprocessing* akan sangat mempengaruhi segi kualitas pada output akhir. Jika data yang diinput tidak berkualitas, maka hasil deteksi juga tidak berkualitas dan diragukan keakuratannya. Dalam proses

JATIMIKA
Jurnal Kreativitas Mahasiswa Informatika
Volume 1 Nomor 2 Tahun 2020 Page 99 - 103.
p-ISSN: 2797-6327 e-ISSN:

selanjutnya yaitu *clustering*, dibutuhkan data preprocessing dari output file NS2 berupa file. tracefile (tr). Pada penelitian ini dilakukan proses *preprocessing* data *transformation*, yaitu merubah suatu data agar data yang diperoleh lebih berkualitas. Dengan adanya preprocessing akan meningkatkan hasil analisis yang dilakukan. Tujuan *preprocessing* output file NS2 ini yaitu sebagai inputan *clustering* pada algoritma DBSCAN.

4. Algoritma *Density Based Spatial Clustering of Application with Noise* (DBSCAN)

Algoritma *Density-based Spatial Clustering of Application with Noise* (DBSCAN) merupakan metode clustering yang berbasis kepadatan (*density-based*) dari posisi amatan data dengan prinsip mengelompokkan data yang relatif berdekatan. DBSCAN sering diterapkan pada data yang banyak mengandung noise, hal ini dikarenakan DBSCAN tidak akan memasukkan data yang dianggap noise kedalam cluster manapun. Ide dasar dari DBSCAN adalah memanfaatkan jumlah titik minimal yang harus dimiliki untuk menentukan suatu poin untuk digolongkan menjadi *core point, border point,* atau *noise point* yang disebut MinPts. Selain itu, juga menggunakan threshold yang harus dipenuhi untuk menentukan titik tersebut menjadi *core point, border point,* atau *noise point* yang disebut juga Eps.

Kelebihan dari DBSCAN adalah:

1. Tidak seperti **K-means**, DBSCAN tidak meminta kepada user untuk memasukkan nilai berapa banyak cluster yang akan dibuat
2. DBSCAN dapat membuat cluster dengan beragam bentuk, tidak harus berbentuk circle (lingkaran).
3. DBSCAN dapat membedakan data mana yang outliers.
4. Baik untuk data dalam jumlah besar.
5. Dapat menangani noise.

5. Algoritma Naïve Bayes

Naive bayesian klasifikasi adalah suatu klasifikasi berpeluang sederhana berdasarkan aplikasi teorema *Bayes* dengan asumsi antar variabel penjelas saling bebas (independen). Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu kejadian tertentu dari suatu kelompok tidak berhubungan dengan kehadiran atau ketiadaan dari kejadian lainnya.

Naive Bayes merupakan salah satu metoda machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naïve bayes* proses klasifikasi statistik yang bisa digunakan dalam melakukan prediksi suatu probabilitas pada keanggotaan sebuah class. *Naive Bayesian* dapat digunakan untuk berbagai macam keperluan antara lain untuk klasifikasi dokumen, deteksi spam atau filtering spam, dan masalah klasifikasi lainnya. Dalam hal ini lebih disorot mengenai penggunaan teorema *Naive Bayesian* untuk spam filtering

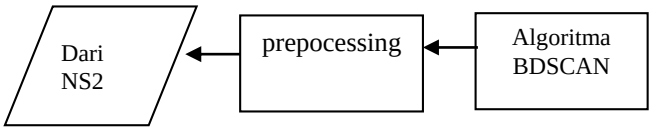
Kelebihan *Naive Bayesian* adalah:

1. Menangani kuantitatif dan data diskrit.
2. Kokoh untuk titik noise yang diisolasi, misalkan titik yang dirata – ratakan ketika mengestimasi peluang bersyarat data.
3. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata – rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
4. Cepat dan efisiensi ruang.
5. Kokoh terhadap atribut yang tidak relevan.

III. HASIL DAN PEMBAHASAN

1. Deskripsi Sistem

Pada perancangan sistem menjelaskan mengenai alur dari proses yang dikerjakan pada tugas akhir kecerdasan buatan ini. Penjelasan yang ada meliputi alur deteksi anomali dan hal-hal yang terkait untuk system deteksi anomali.



Gambar 3. 1 Alur Deteksi Anomali Trafik

2. Dataset NS2 .tracefile

Dalam penelitian tugas akhir mata kuliah kecerdasan buatan ini membutuhkan dataset sebagai data yang diolah pada metode deteksi. Dataset ini merupakan data real time traffic yang belum di ketahui apakah terdapat anomaly didalamnya.

```
+ 14.506886 10 7 udp 512 ------- 1 10.0 1.3 229 3571
14.506886 10 7 udp 512 ------- 1 10.0 1.3 229 3571
r 14.508302 2 0 udp 512 1 14.0 0.1 224 3512
r 14.508302 2 1 udp 512 1 13.0 1.1 224 3513
r 14.508302 3 0 udp 512 1 9.0 0.3 224 3514
r 14.508302 3 1 udp 512 1 10.0 1.3 224 3515
r 14.512915 15 6 tcp 1040 ------- 0 15.0 0.0 310 3552
+ 14.512915 6 2 tcp 1040 0 15.0 0.0 310 3552
14.512915 6 2 tcp 1040 ------ 0 15.0 0.0 310 3552
```

Gambar 3. 2 isi .tracefile

Berdasarkan raw data dengan format diatas, akan dilakukan tahap preprocessing. Tahap preprocessing ini diperlukan untuk mendapatkan fitur – fitur yang relevan

JATIMIKA
Jurnal Kreativitas Mahasiswa Informatika
Volume 1 Nomor 2 Tahun 2020 Page 99 - 103.
p-ISSN: 2797-6327 e-ISSN:

dan akan digunakan sebagai inputan dari algoritma deteksi.

3. Pre-processing

Pada tahap preprocessing, *raw data* hasil dari file .tr yang merupakan output jadi pertama kita akan masukan output yang berasal dari .tr akan menjadi skenario di NS2 dan nantinya menjadi inputan. Preprocessing pada penelitian ini mengacu pada fitur dalam dataset KDDCUP 1999. Pada dataset KDDCUP, terdapat 41 fitur yang dimiliki dalam satu paket, sedangkan yang digunakan pada penelitian ini adalah 4nfitur. trafik normal pada KDDCUP 99 memiliki 11 fitur paling relevan dalam menentukan sebuah trafik tergolong pada trafik normal. 4 fitur ini mengacu pada fitur dengan tingkat kecendrungan paling tinggi pada sebuah trafik normal.

Tabel 3. 1 Fitur Dataset KDD99

Name	*Description*	*Type*
Dur	Duration of the Connection	Continuous
Dest	sent from source to destination	Continuous
Svr rate	% of connections to different hosts	Continuous
Dst count	Count of connections having the same destination host	Continuous

Pengambilan 4 fitur ini dengan proses preprocessing dengan inputan dari hasil output NS2. Berikut merupakan algoritma preprocessing yang dilakukan:

Tabel 3. 2 Algoritma Preprocessing

Algoritma 1 : *Preprocessing* fitur *Duration*

1: Input *out.tr*
2: ***if*** *src_node(x) = src_node(y)* ***and*** *event = "+"* ***then***
3: *time1 ← time;*
4: ***endif***
5: ***if*** *src_node(x) = src_node(y)* ***and*** *event = "r"* ***or*** *"d"* ***then***
6: *time2 ← time;*
7: ***endif***
8: ***if*** *time2 - time1 <> "0"* **then**
9: ***print*** *time2 – time1*
10: ***endif***
11: **end**
12: **until** semua data di *out.tr*

Algoritma 2 : *Preprocessing* fitur *Destination Bytes*

1: Input *out.tr*
2: ***if*** *src_addr(x) = src_addr(y)* ***and*** *dest_addr(x) = dest_addr(y)* ***then***
3: *destbytes ← packet_size;*
4: ***else***
5: *destbytes ← packet_size*
6: ***endif***
7: **end**
8: **until** semua data di *out.tr*

Algoritma 3 : *Preprocessing* fitur *srv_diff_host_rate*

1: Input *out.tr*
2: ***if*** *src_addr(x) = src_addr(y)* ***then***
3: *count ++;*
4: ***if*** *dest_addr(x) = dest_addr(y)* ***then***
5: *diffhost ++*
6: ***endif***
7: ***else***
8: ***(****diffhost/count)*100*
10: ***endif***
11: **end**
12: **until** semua data di *out.tr*

Algoritma 4 : *Preprocessing* fitur *dst_host_count*

1: Input *out.tr*
2: ***if*** *src_addr (x) = src_addr(y)* ***then***
3: *countclient ++;*
4: ***if*** *dest_addr(x) = dest_addr(y)* ***then***
5: *destcountnorm++*
6: ***endif***
7: ***endif***
8: ***if*** *src_addr (x) = src_addr(z)* ***then***
9: *countattack ++;*
10: ***if*** *dest_addr(x) = dest_addr(z)* ***then***
11: *destcountatt++*
12: ***endif***
13: ***endif***
14: **end**
15: **until** semua data di *out.tr*

4. Algoritma DBSCAN

Pada tahap algoritma DBSCAN, data yang dihasilkan algoritma BIRCH yaitu berupa *clustering point* yang akan menjadi inputan. Nilai *Centroid* yang didapatkan pada algoritma sebelumnya yang nantinya akan menjadi acuan utama dari proses algoritma ini dalam meningkatkan kualitas pada hasil cluster. Selanjutnya, akan dilakukan pemberian sebuah label berdasarkan masing masing node yang akan tergolong ke cluster normal ataupun anomali. Dari pelabelan ini akan dibandingkan dengan label prediksi untuk menguji *detection rate* serta akurasi dari hasil cluster

5. Algoritma Naïve Bayes

Naïve Bayes merupakan salah satu metode yang ada di dalam data mining untuk mengklasifikasikan data. metode *Naïve Bayes* memiliki cara kerja dengan menggunakan parameter yang telah ada. Metode ini biasa digunakan dalam statiska untuk menghitung suatu peluang, Bayes Optimal Classifier biasa menghitung

JATIMIKA
Jurnal Kreativitas Mahasiswa Informatika
Volume 1 Nomor 2 Tahun 2020 Page 99 - 103.
p-ISSN: 2797-6327 e-ISSN:

peluang dari satu kelas dari masing – masing kelompok atribut yang ada dan menentukan kelas mana yang paling optimal. Proses pengelompokan atau klasifikasi ini dibagi menjadi dua fase yaitu *learning/training* dan *testing/classify* Pada fase *learning*, sebagian data yang telah diketahui kelas, datanya diumpankan untuk membentuk model perkiraan. Kemudian pada fase *testing*, model yang sudah terbentuk diuji dengan sebagian data. yang nantinya data akan menguji detection rate serta akurasi dari hasil data.

6. Pengujian Accuracy dan Detection Rate

Pada Pengujian kita akan membandingkan metode mana yang lebih baik. Pada suatu pengujian Validasi cluster sangat dibutuhkan untuk menguji seberapa baik hasil clusterisasi yang dihasilkan. Dalam penelitian ini menggunakan *confusion matrix* untuk menguji tingkat *Accuracy* dan *Detection Rate* yang telah dilakukan. Perbandingan data hasil akan dilakukan dengan data prediksi sesuai dengan pelabelan apakah trafik termasuk normal ataukah anomali. Data total yaitu 90.000 data akan dibagi ke 3 bagian dan akan diuji *Detection Rate* dan *Accuracy nya.*

Metode DBSCAN, Berdasarkan data yang ada, tingkat akurasi paling tinggi pada 30.000 data kedua dengan jumlah nilai TP (*True Positive*) mencapai 27.427 dan ternyata memberikan hasil *accuracy* dan *detection rate* hampir mendekati sempurna. Sementara pada 30.000 data pertama dengan selisih antara detection rate dan accuracy nya hampir mencapai 1.5% dengan kondisi nilai TP pada 30.000 data pertama berjumlah 15.365 dan nilai FN (*False Negative*) berjumlah 14.635. Berdasarkan data grafik diatas serta tabel hasil deteksi sebelumnya, semakin kecil nilai FP (*False Positive*) maka akan meningkatkan nilai *detection rate* dan *accuracy*, namun jika perbandingan antara nila FP dan FN hampir sama, maka akan terlihat juga perbandingan antara nilai *detection rate* dan *accuracy* seperti pada 30.000 data pertama dengan selisih nilanya hampir mendekati 1.5

Metode *Naïve Bayes*, Hasil dari data *Naïve Bayes* melalui grafik *Distributions* dan *Radviz* menghasilkan grafik HTTP presentase yang masih kurang dengan nilai akurasi 800 frekuensi dengan probabilitas 0.4 dan TCP kedua nilai akurasinya probabilitas 0.1. Sedangkan HTTP memiliki nilai akurasi yang cukup besar yaitu 1900 frekuensi dan nilai probabilitasnya TCP melebihi dari 0.8. Hasil dari grafik radvid dengan IP 192.168.10.232 dengan penyebaran data lebih banyak dibandingkan dengan IP 23.52.171.89 penyebaran datanya lebih sedikit yang artinya metode ini tidak cukup akurat dalam melakukan deteksi anomali trafik.

IV. SIMPULAN

Berdasarkan analisis perbandingan menggunakan *Naïve Bayes* dan DBSCAN 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 DBSCAN memiliki tingkat akurasi pada 30.000 data pertama dengan selisih antara *detection rate* dan *accuracy* nya hampir mencapai 1.5% dengan kondisi nilai TP pada 30.000 data pertama berjumlah 15.365 nilai FN (*False Negative*) berjumlah 14.635. sedangkan pada data kedua tingkat akurasi yang memiliki akurasi paling tinggi pada contoh 30.000 data kedua dengan jumlah nilai TP (*True Positive*) mencapai 27.427. maka bisa disimpulkan bahwa akurasi metode DBSCAN lebih baik dibandingkan metode Naïve Bayes dalam masalah deteksi anomali trafik.

DAFTAR PUSTAKA

H. Harianto, A. Sunyoto, and S. Sudarmawan, "Optimasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Anomaly dengan Univariate Fitur Selection," Edumatic J. Pendidik. Inform., vol. 4, no. 2, pp. 40–49, 2020, doi: 10.29408/edumatic. v4i2.2433.

[S. Anwar, F. Septian, and R. D. Septiana, "Klasifikasi Anomali Intrusion Detection System (IDS) Menggunakan Algoritma Naïve Bayes Classifier dan Correlation-Based Feature Selection," J. Teknol. Sist. Inf. dan Apl., vol. 2, no. 4, p. 135, 2019, doi: 10.32493/jtsi.v2i4.3453.

D. Yunita, P. Rosyani, and R. Amalia, "Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes," J. Inform. Univ. Pamulang, vol. 3, no. 4, p. 209, 2018, doi: 10.32493/informatika. v3i4.2032.

I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm)," Ilk. J. Ilm., vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom. v11i1.361.17-24.

A. R. Aritonang, Sutarman, and P. Sihombing, "Analisis Subspace Clustering Menggunakan DBSCAN dan SUBCLU Untuk Proyeksi Pekerjaan Alumni Perguruan Tinggi," Teknovasi, vol. 02, pp. 33–60, 2015.

M. A. Shauma, Y. Purwanto, and A. Novianty, "Deteksi Anomali Trafik Menggunakan Algoritma Birch Dan Dbscan Pada Streaming Traffic," eProceedings Eng., vol. 3, no. 3, pp. 5004–5012, 2016, [Online]. Available: https://libraryeproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/3132.