Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 2 ; Juli 2025 ; Page 245-252
E-ISSN : 2615-5133 ; P-ISSN : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Analisis Komperatif Akurasi Deteksi Phising Pada Jaringan Komputer Dengan Metode *Naïve Bayes* Dan *Support Vector Machine* (SVM)

Muit Sunjaya[1]
[1]Sistem Informasi Teknologi Informasi, STMIK Triguna Dharma
Email: [1]muitsunjaya04@gmail.com
Email Penulis Korespondensi: muitsunjaya04@gmail.com

Article History:
Received Jun 15th, 2025
Revised Jun 30th, 2025
Accepted Jul 30th, 2025

Abstrak

Serangan phishing merupakan salah satu bentuk ancaman siber yang bertujuan untuk mencuri data sensitif dengan cara menipu pengguna melalui jaringan komputer. Untuk mendeteksi dan mencegah serangan semacam ini, diperlukan sistem yang mampu mengenali pola ancaman secara cerdas, salah satunya adalah Intrusion Detection System (IDS). IDS dapat bekerja berdasarkan dua pendekatan utama, yaitu rule based (signature based) dan behavior based. Penelitian ini menggunakan pendekatan behavior based yang memanfaatkan dataset serta algoritma klasifikasi untuk mendeteksi serangan phishing. Tidak semua algoritma data mining memiliki kinerja klasifikasi yang optimal, oleh karena itu dilakukan analisis komparatif terhadap algoritma Naive Bayes dan Support Vector Machine (SVM) dengan kernel Linear, Polynomial, dan Sigmoid. Dataset yang digunakan berasal dari UCI Machine Learning Repository dengan fokus pada deteksi phishing. Evaluasi dilakukan berdasarkan metrik confusion matrix, precision, recall, dan f1-score. Hasil pengujian menunjukkan bahwa algoritma SVM Polynomial memiliki akurasi tertinggi, sedangkan Naive Bayes menunjukkan akurasi terendah. Hasil ini menunjukkan bahwa pemilihan algoritma sangat berpengaruh terhadap performa sistem deteksi phishing.

Kata kunci: *Phishing, IDS, Klasifikasi, Naive Bayes, Support Vector Machine (SVM)*

Abstract

Phishing attacks are a type of cyber threat aimed at stealing sensitive data by deceiving users through computer networks. To detect and prevent such attacks, intelligent systems are required to recognize threat patterns, such as an Intrusion Detection System (IDS). IDS operates based on two main approaches: rule-based (signature-based) and behavior-based. This study employs the behavior-based approach, utilizing datasetsand classification algorithms to detect phishing attacks. Since not all data mining algorithms deliver optimal classification performance, this research conducts a comparative analysis ofthe Naive Bayes algorithm and Support Vector Machine (SVM) with Linear, Polynomial, and Sigmoid kernels. The dataset used is sourced from the UCI Machine Learning Repository, focusing on phishing detection. The evaluation is carried outusing confusion matrix metrics, precision, recall, and F1-score. The results show that the SVM with Polynomial kernel achieved the highest accuracy, while Naive Bayes recorded the lowest. These findings indicate that the choice of algorithm significantly influences the performance of phishing detection systems.

Keywords: *Phishing, IDS, Classification, Naive Bayes, Support Vector Machine (SVM)*

1. PENDAHULUAN

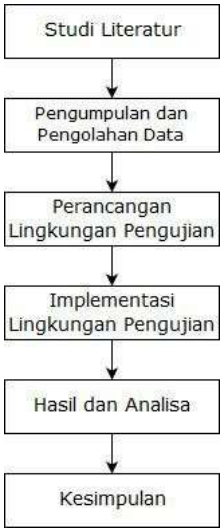
Perlindungan jaringan menjadi aspek krusial dalam pengelolaan sistem informasi, terutama di tengah pesatnya adopsi teknologi digital. Semakin bertambahnya jumlah pengguna internet, meningkatnya volume data, serta kompleksitas arsitektur jaringan turut membuka celah yang lebih besar bagi serangan siber. Salah satu jenis ancaman yang paling sering dijumpai adalah phishing**,** yaitu teknik rekayasa sosial yang menipu pengguna agar secara tidak sadar memberikan informasi sensitif seperti akun, sandi, dan data finansial.[1] Phishing seringkali tidak terdeteksi secara langsung oleh pengguna, menjadikannya sebagai bentuk serangan yang sulit dihindari dan berisiko tinggi terhadap keamanan data. Data terkini dari IBM X-Force Threat Intelligence Index (2023) menunjukkan bahwa phishing merupakan teknik serangan awal yang paling dominan, digunakan dalam sekitar 41% insiden pelanggaran keamanan di seluruh dunia. Hal ini

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 2 ; Juli 2025 ; Page 245-252
E-ISSN : 2615-5133 ; P-ISSN : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

diperkuat oleh laporan Verizon Data Breach Investigations Report (2023) yang menyatakan bahwa 74% dari kasus pelanggaran data disebabkan oleh kesalahan manusia, dengan phishing sebagai salah satu faktor utamanya. Temuan-temuan ini menegaskan perlunya sistem yang mampu mengenali dan merespons serangan phishing secara otomatis dengan tingkat akurasi yang tinggi.[2]. Untuk menjawab tantangan tersebut, salah satu pendekatan yang umum digunakan adalah melalui implementasi Intrusion Detection System (IDS). IDS merupakan sistem yang dirancang untuk memantau lalu lintas jaringan dan mendeteksi pola aktivitas yang menyimpang dari perilaku normal.[3] Secara umum, IDS terbagi menjadi dua kategori, yaitu pendekatan berbasis tanda tangan (signature-based) dan berbasis perilaku (behavior-based). Sistem signature-based mendeteksi ancaman berdasarkan pola serangan yang sudah dikenali sebelumnya, sedangkan behavior-based menganalisis lalu lintas jaringan untuk mendeteksi anomali, sehingga memiliki kemampuan lebih baik dalam mengidentifikasi jenis serangan baru yang belum tercatat dalam database.[4] Dalam sistem berbasis perilaku, keberhasilan deteksi sangat dipengaruhi oleh algoritma klasifikasi yang digunakan. Oleh karena itu, pemilihan metode machine learning yang tepat menjadi sangat penting. Algoritma seperti Support Vector Machine (SVM) dan Naive Bayes telah banyak digunakan dalam berbagai penelitian karena keunggulan masing-masing dalam memproses data klasifikasi. SVM dikenal karena kemampuannya dalam memisahkan data non-linear melalui berbagai jenis kernel, sedangkan Naive Bayes menawarkan pendekatan probabilistik yang sederhana namun efisien. Sebuah studi oleh Wu et al. (2022) menyebutkan bahwa algoritma SVM dengan penyesuaian kernel tertentu mampu memberikan performa yang sangat baik dalam mendeteksi serangan berbasis jaringan.[5] Di sisi lain, Naive Bayes masih menunjukkan hasil yang kompetitif terutama dalam skenario dengan jumlah data yang tidak terlalu besar. Akan tetapi, sejumlah penelitian terdahulu umumnya masih menggunakan dataset lama seperti KDD Cup 1999 dan DARPA 1998, yang dianggap kurang relevan dalam menggambarkan karakteristik serangan siber masa kini. Penelitian ini bertujuan untuk melakukan perbandingan tingkat akurasi dalam deteksi serangan phishing menggunakan algoritma Naive Bayes serta SVM dengan tiga jenis kernel yang berbeda: Linear, Polynomial, dan Sigmoid. Dataset yang digunakan adalah ISCX 2012, yang disusun oleh Canadian Institute for Cybersecurity dan dirancang untuk merepresentasikan lalu lintas jaringan modern secara realistis. Tahapan dalam penelitian ini mencakup preprocessing data, seleksi fitur yang relevan, pelatihan model, serta evaluasi performa klasifikasi menggunakan metrik seperti confusion matrix, precision, recall, f1-score, dan ROC curve. Teknik random sampling digunakan untuk membagi data menjadi 60% data pelatihan dan 40% data pengujian. Melalui penelitian ini, diharapkan dapat diperoleh algoritma klasifikasi yang paling optimal dalam mendeteksi serangan phishing, sehingga dapat mendukung pengembangan sistem IDS yang lebih responsif, akurat, dan adaptif terhadap dinamika ancaman di era digital saat ini.[6]

2. METODOLOGI PENELITIAN

Metode Penelitian ini akan dilakukan beberapa tahap, ini akan memberikan gambar Flochart tahapan penelitian



Gambar: Diagram Alir Metodologi Penelitian

2.1Tahapan Penelitian

Adapun teknik yang digunakan dalam pengumpulan data informasi terkait membandingkan akurasi algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mendeteksi serangan phishing pada jaringan komputer :

1. **Perumusan Masalah.**
   Menentukan fokus masalah utama, yaitu bagaimana membandingkan akurasi algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mendeteksi serangan phishing pada jaringan komputer.
2. **Studi Literatur.**

Melakukan kajian terhadap jurnal, buku, dan laporan penelitian sebelumnya yang berkaitan dengan:
   a. Serangan phishing
   b. Sistem Intrusion Detection System (IDS)
   c. Algoritma Naive Bayes dan SVM
   d. Dataset yang digunakan dalam penelitian keamanan jaringan

3. **Pengumpulan Data (Dataset)**
   Menggunakan dataset ISCX 2012 dari Canadian Institute for Cybersecurity yang memuat data lalu lintas jaringan termasuk serangan phishing. Dataset diunduh dalam format .csv untuk diproses lebih lanjut.
4. **Preprocessing Data**
   a. Pembersihan data: Menghapus data kosong, data tidak relevan, atau duplikasi.
   b. Encoding: Mengubah data kategori (non-numerik) menjadi bentuk numerik agar bisa diproses algoritma.
   c. Normalisasi: Menyesuaikan skala nilai fitur agar seragam.
   d. Seleksi fitur**:** Menghapus atribut yang tidak mendukung klasifikasi agar model lebih optimal.
5. **Pembagian Data**
   a. 60% data latih (training data)
   b. 40% data uji (testing data)
      Menggunakan metode random sampling agar hasil tidak bias a.
6. **Implementasi Model**
   a. Naive Bayes
   b. SVM dengan kernel Linear
   c. SVM dengan kernel Polynomial
   d. SVM dengan kernel Sigmoid
      Setiap model dilatih menggunakan data latih dan diuji pada data uji.
7. **Evaluasi Kinerja Model**
   a. Akurasi (Accuracy)
   b. Presisi (Precision)
   c. Recall (Sensitivity)
   d. F1-Score
   e. Confusion Matrix
   f. ROC Curve
8. **Analisis dan Perbandingan**
   Menganalisis hasil pengujian masing-masing algoritma dan membandingkan kinerja berdasarkan nilai evaluasi untuk menentukan model terbaik dalam deteksi phishing.

2.2 Intrusion Detections System

Intrusion Detection System (IDS) merupakan sebuah perangkat keras maupun perangkat lunak yang mampu melakukan suatu deteksi pada sebuah aktivitas mencurigakan yang terjadi pada suatu jaringan komputer maupun sistem komputer yang mencurigakan dalam sebuah sistem atau jaringan.[4] Menurut (Wu, 2009) Intrusion Detection System (IDS) merupakan tindakan untuk melakukan sebuah deteksi dari beberapa traffic paket yang dalam prosesnya tidak sesuai atau tidak diharapkan terjadi pada sebuah jaringan.

IDS dalam kemampuannya melakukan deteksi intrusi ataupun serangan pada sebuah jaringan dapat dikategorikan menjadi 2 kategori yaitu sebagai berikut:

1. *Network-based Intrusion Detection System* (NIDS) yang dalam proses melakukan deteksi jika terdapat intrusi atau serangan, maka akan dilakukan analisis untuk seluruh lalu lintas yang terjadi pada jaringan tersebut. Pada dasarnya sebuah NIDS terdapat pada sebuah segmen penting yang ada pada jaringan, yaitu dapat dikatakan sebagai pintu masuk pada sebuah jaringan. Meskipun demikian kelemahan pada NIDS yaitu sedikit lebih rumit saat dilakukan implementasi pada sebuah jaringan yang menggunakan *switch Ethernet*. Namun saat ini beberapa dari vendor *switch Ethernet* sudah melakukan penerapan fungsi IDS pada switch yang telah dibuatnya supaya dapat memonitor koneksi *port*.
2. *Host-based Intrusion Detection System* (HIDS) merupakan IDS yang dalam proses kerjanya hanya dapat mendeteksi sebuah intrusi hanya pada *host* tempat dimana dilakukan implementasi IDS. HIDS melakukan pengamatan maupun pemantauan segala aktivitas hanya pada sebuah *host* jaringan individual apakah didalamnya terdapat aktivitas mencurigakan atau sebuah percobaan penyusupan. HIDS lebih sering diletakkan pada beberapa *server* penting, diantaranya pada *firewall*, *web server*, atau *server* yang terkoneksi ke internet.

2.2.1 Naïve Bayes

Naive Bayes menurut Xhemali, et al (2009) adalah sebuah metode atau algoritma klasifikasi sederhana (*simple*), yang mampu berkontribusi pada keputusan akhir dan pada setiap atributnya memiliki sifat *independent*. Menurut Hang dkk (2006) *Naive Bayes* merupakan proses klasifikasi statistik yang bisa digunakan dalam melakukan prediksi suatu

probabilitas pada keanggotaan sebuah *class*. Sedangkan menurut Bustami (2013) *Naive Bayes* merupakan metode pengklasifikasian suatu probabilitas dan statistik yang diperoleh Thomas Bayes seorang ilmuwan Inggris dengan cara melakukan prediksi peluang di masa depan berdasarkan pengalaman pada masa sebelumnya.[7]

Beberapa tujuan sistem pendukung keputusan menurut Turban dalam buku [8] yaitu:

a. Membantu manajer dalam pengambilan keputusan dalam permasalahan.
b. Memberikan dukungan atas pertimbangan manajer dengan kata lain tidak mengganti posisi manajer.
c. Meningkatkan efektivitas keputusan yang diambil manajer lebih dari pada perbaikan efisiensinya.
d. Para pengambil keputusan menggunakan komputer untuk melakukan banyak komputasi secara cepat dan biaya yang murah.
e. Peningkatan produktivitas. Dalam membangun satu kelompok pengambil keputusan, terutama para pakar, bisa sangat mahal. Dengan adanya pendukung terkomputerisasi dapat mengurangi kelompok dan memungkinkan para anggotanya berada di berbagai lokasi yang berbeda-beda (menghemat biaya).[2]

2.3 Support Vector Machine *(SVM)*

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, dan Vapnik yang dikemukakan pada tahun 1992 di Annual Workshop on Computational Learning Theory.[9] Konsep dasar yang ada pada SVM merupakan beberapa teori komputasi yang sudah ada puluhan tahun sebelumnya, misalnya saja margin hyperplane.[10] Pada tahun 1950 seorang Aronszajn memperkenalkan konsep kernel SVM serta kosep-konsep pendukung yang lain. Beberapa komponen yang ada tersebut belum terdapat upaya untuk dirangkai hingga tahun 1992. Terdapat beberapa pilihan fungsi Kernel dari SVM diantranya (Suykens, Gestel, Brabanter, Moor, & Vandewalle, 2002):

a. SVM *Linear*
b. SVM *Polynomial*
c. Kernel RBF *(Radial Basis Function)*
d. Kernel MLP *(Multi Layer Perceptron)*
e. *Tangent Hyperbolic (sigmoid)*

2.4 Confusion Matrix

Confusion matrix menurut Han dan Kamber (2011) dapat diartikan sebagai suatu alat yang memiliki fungsi untuk melakukan analisis apakah classifier tersebut baik dalam mengenali tuple dari kelas yang berbeda.[11] Nilai dari True-Positive dan True-Negative memberikan informasi ketika classifier dalam melakukan klasifikasi data bernilai benar, sedangkan False- Positive dan False-Negative memberikan informasi ketika classifier salah dalam melakukan klasifikasi data.

		Predicted class		
		yes	*no*	Total
Actual class	*yes*	*TP*	*FN*	*P*
	no	*FP*	*TN*	*N*
	Total	*P'*	*N'*	*P + N*

Gambar 2. *Confusion Matrix* menampilkan total *positive* dan *negative tuple* (Han dan Kamber, 2011)

TP *(True Positive)* → Jumlah data dengan nilai sebenarnya positif dan nilai prediksi positif
FP *(False Positive)* → Jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif
FN *(False Negative)* → Jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif
TN *(True Negative)* → Jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif

2.5 Kurva ROC *(Receiver Operating Characteristic)*

Kurva ROC menunjukan visualisasi dari akurasi model dan membandingkan perbedaan antar model klasifikasi. Receiver Operating Characteristic (ROC) mengekspresikan confusion matrix (Vercellis, 2009).[12] ROC merupakan grafik dua dimensi dimana false positives sebagai garis horizontal sedangkan true positives untuk mengukur perbedaaan performasi metode yang digunakan. Kurva ROC merupakan teknik untuk memvisualisasi dan menguji kinerja pengklasifikasian berdasarkan performanya (Gorunescu, 2011). Model klasifikasi yang lebih baik adalah yang mempunyai kurva ROC lebih besar (Vercellis, 2009).

2.6 Dataset ISCX 2020

Dataset Information Security Center of eXcellence (ISCX) merupakan data yang dikembangkan oleh Fakultas Ilmu Komputer, Universitas New Brunswick dari tahun 2009 sampai tahun 2012.[13] Penelitian yang dilakukan oleh (Shiravi, Tavallaee dan Ghorbani, 2018) menjelaskan tentang ISCX dan jenis pendekatan yang digunakan dalam mengembangkan dataset. Seluruh dataset berlabel ISCX terdiri dari 157867 paket dengan 19 features dan mengumpulkan lebih dari tujuh hari aktivitas jaringan (yaitu normal dan intrusi). Berikut kumpulan data simulasi ISCX berdasarkan skenario serangan:

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 2 ; Juli 2025 ; Page 245-252
E-ISSN : 2615-5133 ; P-ISSN : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

1. Infiltrasi jaringan dari dalam
2. HTTP *Denial of Service*
3. *Distributed DoS* menggunakan IRC*Botnet*
4. *Brute-force SSH*

Berikut features pada Intrusion Detection System dataset ISCX 2012:

Tabel 1 Daftar *Features* pada Dataset ISCX

No.	Nama	No.	Nama
1	*appName*	11	*Src TCPFlagsDescription*
2	*totalSourceBytes*	12	*Dst TCPFlagsDescription*
3	*totalDestinationBytes*	13	*Source*
4	*totalDestinationPacket*	14	*protocolName*
5	*totalSourcePacket*	15	*sourcePort*
6	*Src PayloadAsBase64*	16	*Destination*
7	*Src PayloadAsBaseUTF*	17	*destinationPort*
8	*Dst PayloadAsBase64*	18	*startDateTime*
9	*Dst PayloadAsBaseUTF*	19	*stopDateTime*
10	*direction*		

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi hasil dan pembahasan dari topik penelitian, yang bisa dibuat terlebih dahulu metodologi penelitian. Bagian ini juga mempresentasikan penjelasan yang berupa penjelasan, gambar,tabel dan lainnya. (Zanakis)

3.1 Hasil

Hasil dari lingkungan pengujian pada proses klasifikasi dengan menggunakan metode Naive Bayes, SVM Linear, SVM Polynomial, serta SVM Sigmoid menampilkan persentase nilai accuracy, precision, recall, dan f1 score dari masing-masing classifier.

a) Confusion Matrix

Dari hasil perhitungan *confusion matrix* yang dilakukan pada proses klasifikasi pada modul *classify* dengan metode *Naive Bayes*, maka dihasilkan ringkasan nilai sebagai berikut:

Tabel 2 Perhitungan *Confusion Matrix (Naive Bayes)*

Jenis Paket	Jumlah Paket	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
TN	0				
FP	0				
FN	5978	85,055	100	85,055	91,924
TP	34022				

Dari hasil perhitungan *confusion matrix* yang dilakukan pada proses klasifikasi pada modul *classify* dengan metode SVM *Linear*, maka dihasilkan ringkasan nilai sebagai berikut:

Tabel 3 Perhitungan *Confusion Matrix* (SVM *Linear*)

Jenis Paket	Jumlah Paket	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
TN	0				
FP	1				
FN	0	99,995	99,995	100	99,997
TP	1999				

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 2 ; Juli 2025 ; Page 245-252
E-ISSN : 2615-5133 ; P-ISSN : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

9

Dari hasil perhitungan *confusion matrix* yang dilakukan pada proses klasifikasi pada modul *classify* dengan metode SVM *Polynomial*, maka dihasilkan ringkasan nilai sebagai berikut:

Tabel 4 Perhitungan *Confusion Matrix* (SVM *Polynomial*)

Jenis Paket	Jumlah Paket	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
TN FP FN TP	0 1 1 19998	99,999	99,994	99,994	99,994

Dari hasil perhitungan confusion matrix yang dilakukan pada proses klasifikasi pada modul classify dengan metode SVM Sigmoid, maka dihasilkan ringkasan nilai sebagai berikut:

Tabel 5 Perhitungan Confusion Matrix (SVM Sigmoid)

Jenis Paket	Jumlah Paket	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
TN FP FN TP	0 1 0 19999	99,995	99,995	100	99,997

b) Running Time

Running time atau yang biasa disebut sebagai waktu komputasi yang dibutuhkan oleh sebuah sistem dalam melakukan penyelesaian masalah dan juga membangun sebuah model pada sebuah komputer. Waktu komputasi dilakukan perhitungan mulai dari algoritma berjalan hingga saat algoritma berhenti. Berdasarkan hasil penelitian, dalam melakukan proses klasifikasi hingga memberikan hasil output, masing-masing metode Naive Bayes, SVM Linear, SVM Polynomial, dan SVM Sigmoid membutuhkan waktu yang berbeda-beda, dapat dilihat pada tabel berikut:

Tabel 6 Hasil Perbandingan Running Time Classifier

Classifier	***Running Time (seconds)***
Naive Bayes	16.189985990524292
SVM *Linear*	1201.6530289649963
SVM *Polynomial*	652.7967011928558
SVM *Sigmoid*	75.86262702941895

3.2 Pembahasan

Hasil dari perhitungan confusion matrix dengan metode Naive Bayes pada Tabel 5.1 dengan jumlah paket yang terdeteksi sebagai True-Negative sebesar 0 data, False-Positive sebesar 0 data, False-Negative sebesar 5978 data dan True-Positive sebesar 34022 data, menghasilkan nilai accuracy 85,055%, precision 100%, recall 85,055% dan F1 score 91,924%. Hal ini berarti nilai yang dihasilkan dari perbandingan data, atau yang diidentifikasikan apakah benar merupakan *attack* atau data normal dari total keseluruhan data hanya 85,055% yang diidentifikasikan benar. Jika dilihat dari tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan sistem menghasilkan nilai 100% yang berarti sangat tepat. Tetapi jika dilihat dari tingkat perolehan keberhasilan sistem dalam menemukan kembali sebuah informasi hanya sebesar 85,055%. Pada pengujian ini dapat dikatakan bahwa kualitas klasifikasi cukup berhasil karena memperoleh nilai *precision* dan *recall* yang tinggi.

Hasil dari perhitungan confusion matrix dengan metode SVM Linear pada Tabel 5.2 dengan jumlah paket yang terdeteksi sebagai True-Negative sebesar 0 data, False-Postive sebesar 1 data, False-Negative sebesar 0 data dan True-

Positive sebesar 19999 data, menghasilkan nilai accuracy 99,995%, precision 99,995%, recall 100% dan F1 score 99,997%. Hal ini berarti nilai yang dihasilkan dari perbandingan data, atau yang diidentifikasikan apakah benar merupakan attack atau data normal dari total keseluruhan data memiliki persentase yang cukup tinggi sebesar 99,995%. Jika dilihat dari tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan sistem menghasilkan nilai 99,995% yang berarti tingkat ketepatannya tinggi. Tetapi jika dilihat dari tingkat perolehan keberhasilan sistem dalam menemukan kembali sebuah informasi memiliki persentase yang sangat tinggi sebesar 100%. Pada pengujian ini dapat dikatakan bahwa kualitas klasifikasi berhasil karena memperoleh nilai precision dan recall yang sangat tinggi.

Hasil dari perhitungan confusion matrix dengan metode SVM Polynomial pada Tabel 5.3 dengan jumlah paket yang terdeteksi sebagai True-Negative sebesar 0 data, False-Postive sebesar 1 data, False-Negative 1 data, dan True- Positive sebesar 19998 data, menghasilkan nilai accuracy 99,999%, precision 99,994%, recall 99,994% dan F1 score 99,994%. Hal ini berarti nilai yang dihasilkan dari perbandingan data, atau yang diidentifikasikan apakah benar merupakan attack atau data normal dari total keseluruhan data memiliki persentase yang hampir sempurna yaitu 99,999%. Jika dilihat dari tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan system dengan tingkat perolehan keberhasilan sistem dalam menemukan kembali sebuah informasi memiliki persentase yang sama yaitu sebesar 99,994%. Hal ini berarti, pada pengujian yang dilakukan dapat dikatakan bahwa kualitas klasifikasi cukup berhasil karena memperoleh nilai precision dan recall yang tinggi.

Hasil dari perhitungan confusion matrix dengan metode SVM Sigmoid pada Tabel 5.4 dengan jumlah paket yang terdeteksi sebagai True-Negative sebesar 0 data, False-Postive sebesar 1 data, False-Negative sebesar 0 data, dan True-Positive sebesar 19999 data, menghasilkan nilai accuracy 99,995%, precision 99,995%, recall 100% dan F1 score 99,997%. Hal ini berarti nilai yang dihasilkan dari perbandingan data, atau yang diidentifikasikan apakah benar merupakan attack atau data normal dari total keseluruhan data memiliki persentase yang cukup tinggi sebesar 99,995%. Jika dilihat dari tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan sistem menghasilkan nilai 99,995% yang berarti tingkat ketepatannya tinggi. Tetapi jika dilihat dari tingkat perolehan keberhasilan sistem dalam menemukan kembali sebuah informasi memiliki persentase yang sangat tinggi sebesar 100%. Pada pengujian ini dapat dikatakan bahwa kualitas klasifikasi berhasil karena memperoleh nilai precision dan recall yang sangat tinggi.

4. KESIMPULAN

Berdasarkan pada hasil pengujian, yang diperoleh mengenai proses klasifikasi pada beberapa metode classifier Naive Bayes, SVM Linear, SVM Polynomial, dan SVM Sigmoid serta dari pembahasan yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Tahapan klasifikasi serangan menggunakan metode behavior based membutuhkan sebuah dataset dan metode. Dengan melakukan perbandingan pola atau aktivitas yang ada pada sebuah data, kemudian dilakukan klasifikasi dengan sebuah metode dan menghasilkan sebuah model. Dari model yang sudah dibangun tersebut diuji dengan data testing menghasilkan sebuah output untuk melihat akurasi apakah sebuah traffic yang ada dapat dikategorikan sebagai intrusi atau bukan.
2. Mekanisme pengolahan dataset ISCX 2012 dilakukan dalam beberapa tahap, yaitu dataset ISCX dilakukan preprocess dengan mengubah format xml menjadi csv, dan mengubah beberapa data dari string menjadi *integer*, selanjutnya menghilangkan beberapa fitur untuk proses klasifikasi, hingga data siap dimasukkan dalam *classifier*.
3. Fitur yang digunakan dalam proses klasifikasi yaitu total Source Bytes, total Destination Bytes, total Destination Packet, total SourcePacket, direction, Source TCPF lags Description, Destination TCPF lags Description, protocol Name, source Port, Destination, destination Port, start Date Time, dan stop Date Time.
4. Performa yang dihasilkan dari *confusion matrix* pada masing-masing *classifier Naive Bayes*, SVM *Linear*, SVM *Polynomial*, dan SVM *Sigmoid* menghasilkan persentase akurasi berturut-turut sebesar 85,055%, 99,995%, 99,999% dan 99,995%.
5. Performa kinerja klasifikasi yang dihasilkan dari kurva ROC pada *classifier Naive Bayes* yaitu baik, SVM *Linear* lemah, SVM *Polynomial* sangat lemah, dan SVM *Sigmoid* lemah. Sedangkan jika dilihat dari kurva ROC dengan *cross-validation* menunjukkan bahwa *classifier Naive Bayes* yaitu lemah dengan nilai AUC 0,5, SVM *Linear* baik dengan nilai AUC 0,75, SVM *Polynomial* sangat lemah dengan nilai AUC 0,33 dan SVM *Sigmoid* lemah dengan nilai AUC 0,5.

UCAPAN TERIMA KASIH

Terima kasih diucapkan kepada Tuhan yang Maha Esa yang memberikan Rahmat dan karunia sehingga mampu menyelesaikan jurnal ini. Kemudian kepada Kampus Stmik Triguna Dharma terimakasih atas kesempatan dan ilmunya yang telah memberikan kepercayaan menjadi Dosen Tetap dikampus Stmik Triguna Dharma

DAFTAR PUSTAKA

[1] M. F. Ridhwan, I. L. Sardi, dan S. Y. Puspitasari, "Rekomendasi Pemilihan Tempat Usaha Makanan dengan Metode COPRAS di Kecamatan Jambangan," *e-Proceeding of Engineering*, vol. 6, no. 2, hlm. 9491–9503, 2019.

[2] A. Darmawan, N. Kustian, dan W. Rahayu, "IMPLEMENTASI DATA MINING MENGGUNAKAN MODEL SVM UNTUK PREDIKSI KEPUASAN PENGUNJUNG TAMAN TABEBUYA," 2018.
[3] D. Darwis, E. Shintya Pratiwi, A. Ferico, dan O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," 2020.
[4] P. A. Octaviani, Y. Wilandari, dan D. Ispriyanti, "PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG," vol. 3, no. 4, hlm. 811–820, 2014, [Daring]. Tersedia pada: http://ejournal-s1.undip.ac.id/index.php/gaussian
[5] A. C. Sitepu, W. Wanayumini, dan Z. Situmorang, "Analisis Kinerja Support Vector Machine dalam Mengidentifikasi Komentar Perundungan pada Jejaring Sosial," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, hlm. 475, Apr 2021, doi: 10.30865/mib.v5i2.2923.
[6] "1444-1653-1-SM".
[7] A. FakultasIlmuKomputer, "ANALISIS SENTIMEN TERHADAP WACANA POLITIK PADA MEDIA MASA ONLINE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES," 2015.
[8] Kusrini, *Konsep Dan Aplikasi Sistem Pendukung Keputusan*, Andi. Yogyakarta, 2007.
[9] "23520".
[10] A. Muhson, "PENGEMBANGAN MEDIA PEMBELAJARAN BERBASIS TEKNOLOGI INFORMASI," 2010.
[11] R. Nooraeni, A. B. Safiruddin, A. F. Afifah, K. D. Agung, dan N. N. Rosyad, "Analisis Sentimen Publik terhadap Sistem Zonasi Sekolah Menggunakan Data Twitter dengan Metode Naïve Bayes Classification," *Faktor Exacta*, vol. 12, no. 4, hlm. 315, Feb 2020, doi: 10.30998/faktorexacta.v12i4.5205.
[12] P. Cavalin dan L. Oliveira, "Confusion matrix-based building of hierarchical classification," dalam *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2019, hlm. 271–278. doi: 10.1007/978-3-030-13469-3_32.
[13] "DIGITAL BUSINESS AND E-COMMERCE MANAGEMENT."