Computatio: Journal of Computer Science and Information Systems, 9/1 (2025), 42 – 51

PENDETEKSIAN SITUS WEB PHISHING DENGAN METODE KLASIFIKASI MENGGUNAKAN PEMBELAJARAN MESIN ANN DAN SVM

Benevito Kevin S.H
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: benevito.535220222@stu.untar.ac.id

Abstrak

Internet merupakan teknologi yang penting di masa yang serba digital ini. Pengguna dapat dengan mudah melakukan berbagai hal mulai dari mencari informasi hingga melakukan transaksi dengan mengunjungi situs di internet. Dengan kemajuan internet ini, banyak juga pelaku penipuan yang membuat situs untuk menarget orang awam yang tidak jeli atau paham untuk mencuri informasi sensitif mereka seperti data, kata sandi, dan berbagai kredensial privat lainnya. Potensi penerapan pembelajaran mesin untuk validasi keaslian situs menjadi salah satu hal yang bisa diterapkan untuk menjaga pengguna. Pembelajaran mesin dapat digunakan untuk membantu manusia mengklasifikasi dan mendeteksi adanya situs yang berbahaya dan juga untuk mencegah terjadinya penipuan dalam skala besar. Pada penelitian ini algoritma pembelajaran mesin yang digunakan adalah Artificial Neural Network (ANN) dan Support Vector Machine (SVM) yang dilatih dari dataset yang mengambil situs-situs di internet. Dari eksperimen dan tes yang dilakukan, akurasi pembelajaran mesin menghasilkan hasil yang memuaskan dengan akurasi rata-rata ANN mencapai 96% dan SVM mencapai 94%. *Hal ini menunjukan potensi penerapan pembelajaran mesin untuk keamanan dan pendeteksian situs yang berbahaya.*

Kata kunci*—ANN, Keamanan Internet, Klasifikasi, Penipuan, SVM*

Abstract

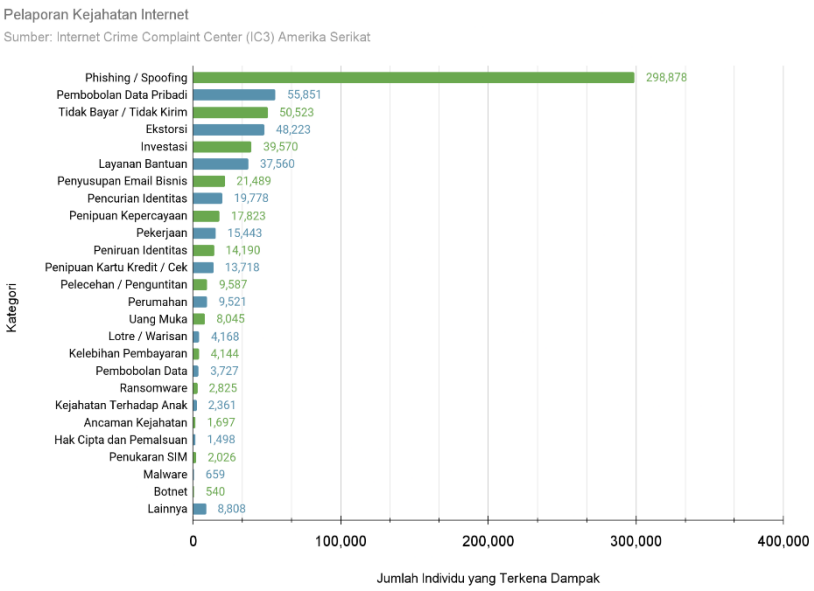
The internet is an important technology in this digital era. Users can easily do various things, from searching for information to making transactions by visiting sites on the internet. With the advancement of the internet, there are also many fraudsters who create sites to target ordinary people who are not observant or knowledgeable to steal their sensitive information such as data, passwords and various other private credentials. The potential for applying machine learning to validate site authenticity is one thing that can be implemented to safeguard users. Machine learning can be used to help humans classify and detect the presence of malicious sites and also to prevent large-scale fraud. In this research, the machine learning algorithms used are Artificial Neural Network (ANN) and Support Vector Machine (SVM) which are drilled from a dataset that takes sites on the internet. From the experiments and tests carried out, the accuracy of the learning machine produced satisfactory results with the average accuracy of ANN reaching 96% and SVM reaching 94%. This shows the potential of applying machine learning for security and malicious site detection.

Keywords—*ANN, Classification, Internet Safety, Fraud, SVM*

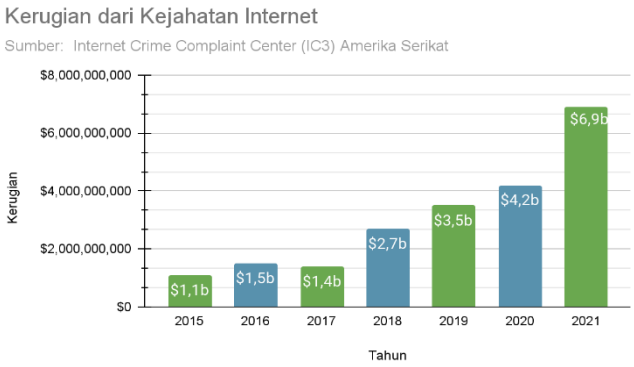
1. PENDAHULUAN

Kemajuan teknologi informasi memberikan kemudahan dalam berbagai aspek kehidupan melalui kehadirannya internet. Mulai dari lingkup perorangan hingga seluruh negara-negara dalam skala global dapat memanfaatkan internet untuk berbagai hal. Melakukan pembayaran dan juga pertukaran informasi yang mudah dan cepat menjadi salah satu hal yang pengguna dapatkan dengan kemajuan internet. Keterbukaan dan kemajuan internet tidak hanya memberikan dampak positif, berbagai pelaku kejahatan dapat menggunakan internet untuk menjalankan operasi mereka.

Penipuan *online* menjadi salah satu hal yang marak terjadi. Statistik pada kejahatan siber menunjukan bahwa *phishing* merupakan salah satu metode penipuan yang paling marak terjadi dengan metode penyerangan yang terus berkembang [1, 2]. Gambar (1) dan (2) menunjukan kasus *phishing* yang dilaporkan dan besarnya kerugian dari kejahatan siber di Amerika Serikat. Kerugian yang mencapai miliaran rupiah mendorong pentingnya metode preventif untuk di buat dan kembangkan [3, 4]. Sistem pendeteksian yang baik dapat menjaga pengguna untuk tidak memberikan informasi kredensial mereka atau mengakses situs-situs yang berbahaya.



Gambar 1 Pelaporan Kejahatan Siber Amerika Serikat



Gambar 2 Kerugian dari Kejahatan Siber Amerika Serikat

Penggunaan pembelajaran mesin untuk mendeteksi situs berbahaya dapat digunakan sebagai salah satu metode preventif. Algoritma pembelajaran yang digunakan di penelitian ini adalah ANN dan SVM. Kedua algoritma akan di tes dan dibandikan untuk mengetahui tingkat akurasi yang didapatkan dengan melakukan klasifikasi pada masalah.

2. STUDI LITERATUR

Berbagai publikasi dan jurnal di beberapa tahun terakhir berfokus pada penggunaan pembelajaran mesin untuk mengembangkan metode deteksi *phishing*. Kemajuan kecerdasan buatan menunjukan potensi untuk keamanan berinternet. Satu publikasi menggunakan *random forest* (RF) dan ANN yang menghasilkan keseluruhan level akurasi 70% - 92.52% [5]. Studi dengan RF, SVM dan ANN menghasilkan akurasi masing-masing 92.369%, 97,451%, dan 97.259% [6]. K-Nearest Neighbour (KNN) dan Decision Tree (DT) juga digunakan untuk mendeteksi situs *phishing* dan menghasilkan tingkat akurasi yang tinggi [7, 8, 9]. Tinjauan sistematis juga dilakukan terhadap metode penelitian yang digunakan di berbagai studi, dengan meninjau 304 studi pembelajaran mesin untuk keamanan siber [10].

Algoritma pembelajaran mesin ANN melakukan klasifikasi dengan melatih input model melalui lapisan-lapisan tersembunyi ke lapisan *output*. Dalam lapisan tersembunyi, informasi dari input diproses melalui serangkaian transformasi matematika kompleks untuk menghasilkan *output* yang sesuai dengan tugas yang diberikan. *Rectified Linear Unit* (ReLU) digunakan pada penelitian ini sebagai fungsi aktivasi ANN. Meskipun kurang optimal dibandingkan dengan metode aktivasi baru, ReLU tetap dapat diandalkan [11, 12]. ReLU dihitung menggunakan persamaan (1), dimana x adalah input ke *neuron*, max(0, x) adalah pembagian hasil, jika x positif atau nol, outputnya akan sama dengan input tersebut, jika input x negatif, outputnya akan menjadi nol, dan f(x) adalah output dari neuron setelah melewati fungsi ReLU. Proses ini juga melibatkan algoritma pembelajaran *backpropagation* untuk mengoptimalkan kinerja jaringan terhadap data pelatihan.

$$f(x) = max(0, x) \tag{1}$$

Algoritma pembelajaran mesin SVM adalah digunakan untuk tugas klasifikasi dan regresi suatu dataset. Cara kerjanya adalah dengan menemukan *hyperplane* terbaik yang memisahkan dua kelas data dalam ruang fitur. *Hyperplane* ini merupakan dimensi yang lebih tinggi dari data untuk membagi margin maksimum antara kelas yang berbeda. Metode SVM sangat efisien untuk mengklasifikasikan pola nonlinier yang dapat dipisahkan menjadi dua kelas [13]. Margin adalah jarak terdekat antara titik data dan *hyperplane* yang terdekat dari kelas yang berbeda. Fungsi kernel untuk *hyperplane* polinomial menggunakan persamaan (2), dimana x dan y adalah vektor input, c adalah konstanta yang ditetapkan, d adalah derajat polinomial dan $x^T y$ adalah *dot product* antara x dan y.

$$K(x, y) = (x^T y + c)^d \tag{2}$$

Model yang digunakan akan dibandingkan dengan menggunakan perpustakaan scikit-learn. Scikit-learn menyediakan fungsi *classification_report,* kurva ROC dan *confusion* matriks dalam modul sklearn.metrics untuk mengukur model klasifikasi. Pengukuran model pembelajaran mesin dilakukan untuk menemukan anomali dan mencetak hasil model [14]. Metrik untuk mengukur model terdiri dari *accuracy*, *precision*, *recall*, dan *f1-score*. Keempat metrik ini dikalkulasi menggunakan persamaan (3) untuk *accuracy*, (4) untuk *precision*, (5) untuk *recall* dan (6) untuk *f1-score*. TP adalah prediksi positif yang benar, TN adalah prediksi negatif yang benar, FN adalah

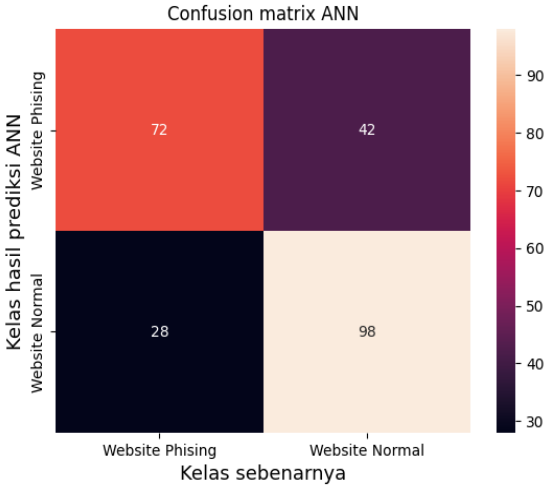
prediksi negatif yang salah dan FP adalah prediksi positif yang salah. Hasil keakuratan model divisualisasikan dengan *confusion* matriks seperti gambar (3) dan kurva ROC seperti gambar (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (3)$$

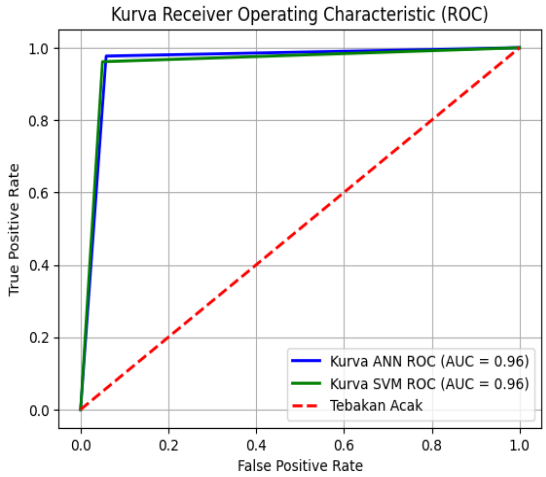
$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad 5)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$



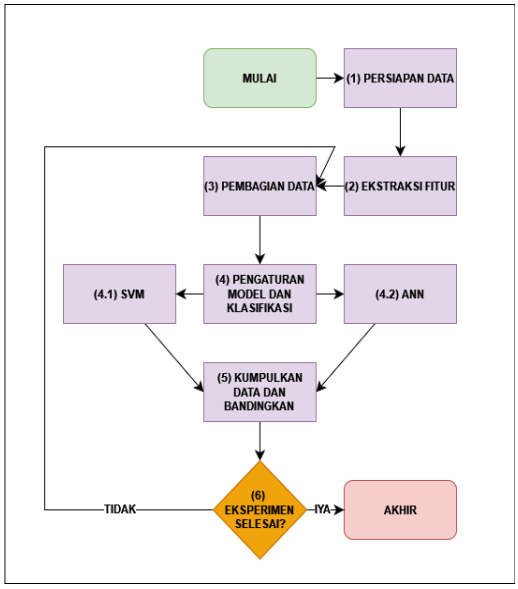
Gambar 3 Confusion Matriks



Gambar 4 Kurva ROC

3. METODE PENELITIAN

Penelitian divisualisasikan dengan menggunakan diagram alir gambar (5).



Gambar 5 Diagram Alir Penelitian

Penjelasan tahapan metode penelitian:

1. Persiapan data
   Dataset dikumpulkan lalu dibersihkan untuk dijadikan bahan pelatihan dan pengujian model.
2. Ekstraksi fitur
   Identifikasi dan ekstraksi fitur yang relevan dari dataset.
3. Pembagian data
   Bagi data menjadi data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi kinerja model.
4. Pengaturan Model dan Klasifikasi
   Lakukan pengaturan model untuk SVM dan ANN dengan mengubah parameter pembelajaran mesin.
   4.1. Latih model SVM.
   4.2. Latih model ANN.
5. Kumpulkan data dan bandingkan
   Kumpulkan hasil yang didapatkan dalam bentuk laporan klasifikasi model dan grafik ROC lalu bandingkan kinerja SVM dan ANN menggunakan skor yang telah di rata-rata.
6. Eksperimen selesai
   Lakukan perbaikan model atau berhenti melakukan eksperimen.

4. HASIL DAN PEMBAHASAN

4.1 Dataset

Dataset yang digunakan di buat oleh Choon Lin Tan yang di unggah di situs Mendeley Data. Dataset ini memiliki 48 fitur yang diekstraksi dari 5000 situs phishing dan 5000 situs normal. 3 fitur menggunakan tipe data float dan 45 fitur menggunakan tipe data integer [15]. Sumber situs phishing diambil dari Phishtank dan OpenPhish. Untuk situs normal diambil dari Alexa dan Common Crawl. Di berbagai publikasi penggunaan dataset yang besar dan sampel yang seimbang mendorong hasil akurasi model [16, 17, 18].

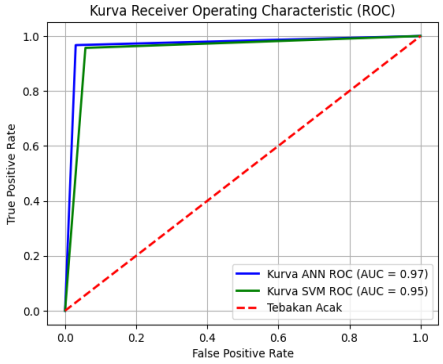
4.2 Pra-pemprosesan data

Tahap pra-pemrosesan dilakukan dengan mengecek kekomplitan seluruh data yang akan digunakan seperti data dengan *missing value* atau tipe data yang *null* sehingga dapat digunakan untuk pelatihan model pembelajaran mesin.

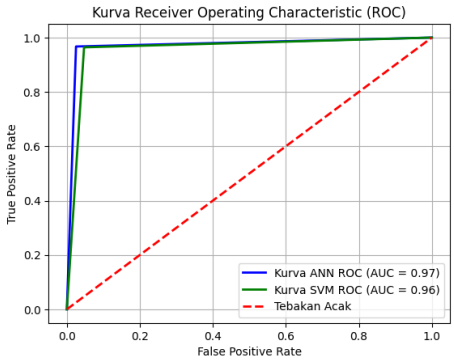
4.3 Skenario eksperimen

Eksperimen dilakukan dengan melakukan dua kali pembagian data dan pengubahan parameter model pada setiap tes. Pembagian data yang pertama adalah 80/20, 80% sampel uji dan 20% sampel tes. Pembagian data yang ke dua adalah 60/40, 60% sampel uji dan 40% sampel tes. Pembagian ini dilakukan untuk mengetahui seberapa besar pengaruh pada perbandingan hasil akhir tes [19, 20, 21]. Model SVM menggunakan *kernel* polinomial untuk seluruh tes yang dilakukan. Parameter model yang diubah adalah derajat polinomial, coef0 dan parameter regulasi C. Model ANN menggunakan aktivasi ReLU untuk seluruh tes. Parameter model ANN yang diubah adalah besar iterasi, banyak lapisan tersembunyi, dan kekuatan regulasi L2 model. Kedua model dites sebanyak 100 kali untuk setiap pembagian data, dengan total tes keseluruhan menjadi 400 uji coba.

4.4 Hasil eksperimen



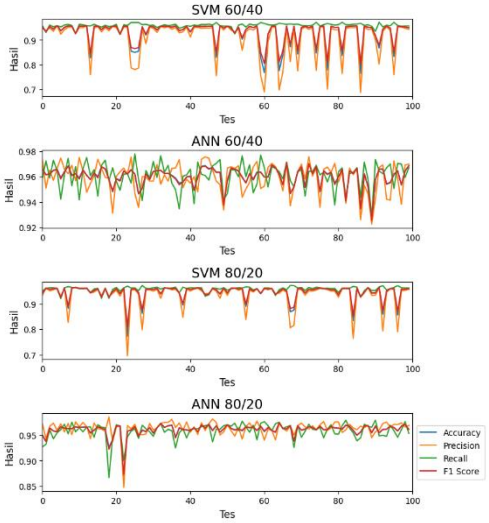
Gambar 6 Kurva ROC 60/40



Gambar 7 Kurva ROC 80/20

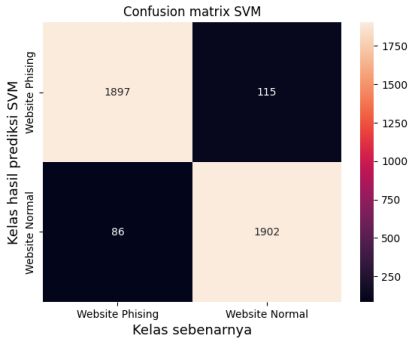
Kurva Receiver Operating Characteristic (ROC) digunakan untuk memvisualisasikan *rate* prediksi oleh model. Kurva ROC memberikan gambaran yang komprehensif tentang seberapa baik model klasifikasi mampu membedakan antara kelas positif dan negatif pada berbagai nilai *threshold*, sehingga membantu dalam mengevaluasi dan membandingkan kinerja model klasifikasi yang berbeda. Gambar (6) adalah kurva ROC yang diambil untuk pembagian data 60/40 dan gambar (7) untuk pembagian data 80/20.

Keseluruhan *Accuracy*, *Precision*, *Recall*, dan *F1-score* model yang di tes divisualisasikan menggunakan grafik garis pada gambar (8). Pembagian data dengan 60% data latih dan 40% data uji untuk dua grafik pertama dan pembagian data dengan 80% data latih dan 20% data uji untuk dua grafik terakhir.

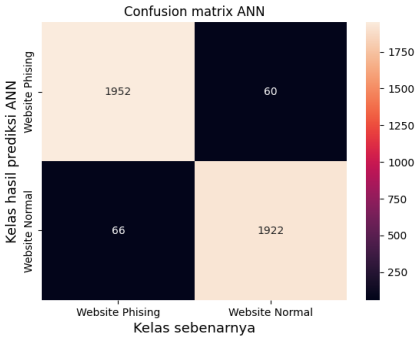


Gambar 8 Visualisasi Hasil Seluruh Tes

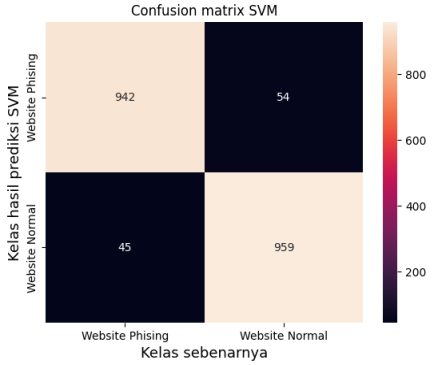
Confusion matriks yang digunakan adalah salah satu hasil dari tes. Gambar (9) menunjukan hasil akurasi SVM dan gambar (10) menunjukan hasil akurasi ANN dari hasil pembagian data 60% latih dan 40% tes. Gambar (11) menunjukan hasil akurasi SVM dan gambar (12) menunjukan hasil akurasi ANN dari hasil pembagian data 80% latih dan 20% uji. Perbandingan dua metode pembelajaran mesin dengan *confusion* matriks menunjukan lebih dalam akurasi dari model yang digunakan.



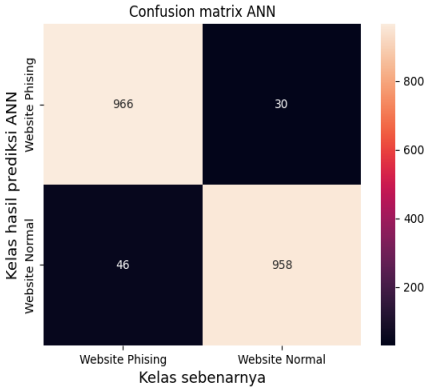
Gambar 9 Confusion Matriks SVM 60/40



Gambar 10 Confusion Matriks ANN 60/40



Gambar 11 Confusion Matriks SVM 80/20



Gambar 12 Confusion Matriks ANN 80/20

4.5 Hasil evaluasi

Tabel 1 Rekapitulasi Hasil Tes

Model	Pembagian Data	Accuracy	Precision	Recall	F1-score
SVM	60/40	0.928772	0.910789	0.959004	0.932496
	80/20	0.944940	0.936221	0.959143	0.946756
ANN	60/40	0.960198	0.959286	0.960855	0.960002
	80/20	0.961260	0.963977	0.958904	0.961313

Tabel (1) menunjukan hasil rata-rata keseluruhan tes yang dilakukan.

4.6 Analisis hasil

Rata-rata akurasi yang didapat oleh kedua model adalah 0.9 untuk seluruh tes yang dilakukan. Hasil akurasi tertinggi yang di dapat adalah 0.97 untuk ANN dan 0.96 untuk SVM. F1-score sebesar 0.9 menunjukkan bahwa kedua algoritma memiliki kinerja yang sangat baik dalam melakukan klasifikasi dan menunjukkan bahwa ANN dan SVM mampu dengan baik mengatasi pertukaran antara *precision* dan *recall*. Meskipun keduanya memiliki kinerja yang tinggi, ANN membutuhkan lebih banyak sumber daya dan komputasi untuk melatih modelnya dibandingkan dengan SVM. SVM cenderung lebih cepat melakukan komputasi tetapi jika

menggunakan parameter derajat polinomial yang tinggi, waktu pelatihan model akan semakin tinggi.

Pembagian data dengan penggunaan lebih banyak data latih menunjukan hasil yang lebih baik pada algoritma SVM. Dengan dataset yang digunakan, ANN tidak menunjukan hasil yang signifikan pada pembagian data 80/20 walaupun dengan waktu komputasi yang lebih lama dibandingkan 60/40. Perbedaan kedua algoritma ini menunjukan bahwa ANN dapat menangani dataset yang kompleks dengan data latih lebih sedikit untuk mendapatkan hasil yang setara atau lebih baik dari algoritma SVM.

5. KESIMPULAN

Kesimpulan penelitian ini adalah pembelajaran mesin dengan algoritma ANN mendapatkan hasil yang lebih baik dari SVM. ANN dapat melakukan prediksi yang lebih akurat yang ditunjukan dari berbagai metrik tes untuk klasifikasi situs phishing. Berikut adalah perbandingan beberapa hal yang di dijadikan perhitungan:

1. Performa:
   - SVM: Pada kedua skenario pembagian data, SVM mendapatkan skor akurasi dan F1 score yang tinggi, dengan nilai rata-rata 0.93 untuk pembagian data 60/40 dan 0.94 untuk 80/20
   - ANN: Dalam skenario pembagian data 60/40 dan 80/20, ANN mendapatkan nilai yang jauh lebih tinggi dibandingkan dengan SVM dengan nilai rata-rata skor akurasi dan F1 score keduanya mencapai 0.96.
2. Pengaruh Pembagian Data:
   - Terlihat bahwa SVM memiliki kinerja yang lebih baik saat menggunakan pembagian data 80/20 dibandingkan dengan 60/40, dengan peningkatan yang signifikan terutama pada SVM. ANN tidak menunjukan perbedaan yang signifikan.
3. Waktu Komputasi
   - Waktu komputasi untuk SVM lebih cepat dibandingkan dengan komputasi ANN. Algoritma ANN sangat berpengaruh pada banyaknya iterasi yang dilakukan dan juga lapisan tersembunyi model, hal ini membuat waktu komputasi lebih lama dibandingkan dengan SVM *hyperplane*.
4. Kompleksitas Dataset:
   - Pada skenario 80/20, ANN berhasil mengungguli SVM dalam semua metrik evaluasi, hal ini menunjukkan keunggulan ANN dalam menangani data yang lebih kompleks dengan menangani jumlah fitur dan data yang banyak.

Pengembangan penelitian di masa depan akan meliputi lebih banyak parameter tes untuk pembelajaran mesin. Pada pembelajaran mesin SVM, kernel yang akan di tes adalah linear, rbf dan sigmoid. Pada pembelajaran mesin ANN, fungsi aktivasi yang akan di tes adalah tanh, *identity*, dan *logistic*. Jumlah tes yang dilakukan juga akan diperbesar.

UCAPAN TERIMA KASIH

Penulis berterima kasih kepada Choon Lin Tan yang membuat dataset pada penelitian ini. Penulis juga berterima kasih kepada dosen Universitas Tarumanagara, Teny Handhayani, yang memberikan masukan penting untuk penelitian ini.

DAFTAR PUSTAKA

[1] Alkhalil, Zainab, et al. "Phishing attacks: A recent comprehensive study and a new anatomy." *Frontiers in Computer Science* 3 (2021).

[2] Alabdan, Rana. "Phishing attacks survey: Types, vectors, and technical approaches." *Future internet* 12.10 (2020).

[3] Cross, Cassandra, and Rosalie Gillett. "Exploiting trust for financial gain: An overview of business email compromise (BEC) fraud." *Journal of Financial Crime* 27.3 (2020).

[4] Sharif, Md Haris Uddin, and Mehmood Ali Mohammed. "A literature review of financial losses statistics for cyber security and future trend." *World Journal of Advanced Research and Reviews* 15.1 (2022).

[5] Mridha, Krishna, et al. "Phishing URL classification analysis using ANN algorithm." *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2021.

[6] Sindhu, Smita, et al. "Phishing detection using random forest, SVM and neural network with backpropagation." *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. IEEE, 2020.

[7] Assegie, Tsehay Admassu. "K-nearest neighbor based URL identification model for phishing attack detection." *Indian Journal of Artificial Intelligence and Neural Networking* 1.2 (2021).

[8] Ojewumi, Theresa O., et al. "Performance evaluation of machine learning tools for detection of phishing attacks on web pages." *Scientific African* 16 (2022).

[9] Karim, Abdul, et al. "Phishing detection system through hybrid machine learning based on URL." *IEEE Access* 11 (2023).

[10] Pruemmer, Julia, Tommy van Steen, and Bibi van den Berg. "A systematic review of current cybersecurity training methods." *Computers & Security* (2023).

[11] Bingham, Garrett, and Risto Miikkulainen. "Discovering parametric activation functions." *Neural Networks* 148 (2022).

[12] Wang, Xueliang, Honge Ren, and Achuan Wang. "Smish: A novel activation function for deep learning methods." *Electronics* 11.4 (2022).

[13] Montesinos López, Osval Antonio, Abelardo Montesinos López, and Jose Crossa. "Support vector machines and support vector regression." *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Cham: Springer International Publishing, 2022.

[14] Kwon, Hyun, et al. "Classification score approach for detecting adversarial example in deep neural network." *Multimedia Tools and Applications* 80 (2021).

[15] Tan, Choon Lin. "Phishing dataset for machine learning: Feature evaluation." *Mendeley Data* 1 (2018).

[16] Ul Hassan, Ietezaz, et al. "Significance of machine learning for detection of malicious websites on an unbalanced dataset." *Digital* 2.4 (2022).

[17] Rajput, Daniyal, Wei-Jen Wang, and Chun-Chuan Chen. "Evaluation of a decided sample size in machine learning applications." *BMC bioinformatics* 24.1 (2023).

[18] Althnian, Alhanoof, et al. "Impact of dataset size on classification performance: an empirical evaluation in the medical domain." *Applied Sciences* 11.2 (2021).

[19] Ch, Rupa, et al. "Computational system to classify cyber crime offenses using machine learning." *Sustainability* 12.10 (2020).

[20] Rácz, Anita, Dávid Bajusz, and Károly Héberger. "Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification." *Molecules* 26.4 (2021).

[21] Salazar, Jose J., et al. "Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy." *Journal of Petroleum Science and Engineering* 209 (2022).