PERBANDINGAN KLASIFIKASI ALGORITMA K-NN, NEURAL NETWORK, NAÏVE BAYES, C 4.5 UNTUK MENDETEKSI WEB PHISING

Eza Nanda [1)], Istikomah [2)],Nurindah A.Amari[3)],Yoga Pristyanto[4)]
[1,2,3]Informatika, Universitas Amikom Yogyakarta
[4] Sistem Informasi, Universitas Amikom Yogyakarta

e-mail: [1]eza.02@students.amikom.ac.id, [2]istikomah.0825@students.amikom.ac.id,
[3]nurindah.amari@students.amikom.ac.id, [4]*yoga.pristyanto@amikom.ac.id*

Abstract

The increasing number of internet users in the world and the rise of web phishing. Based on the Wearesocial report, there are several facts which include the number of world internet users that have reached 4.021 billion people. explained that of the hundreds of millions of internet users in Indonesia, 60 percent have accessed the internet using smartphones. Internet usage is dominated by socializing activities in cyberspace. Evidenced by the large number of world social media users, reaching 3,196 billion users. But with the rise of internet users and social media accompanied by the proliferation of phishing webs. The purpose of this attack is to make users believe that they interact with the official virtual or online site in question. Generally the information sought is in the form of a password, account or victim's credit card number. by way of being directed to a fake web or sending an email, banner or pop-up where the victim is asked to give his personal information. In this study C.4.5, Neural Network and K-Nerest Neighbor algorithm comparison will be conducted to determine the performance of the algorithm in detecting webphising by using classification techniques. Based on the model testing based on the Naïve Bayes method, Decision Tree C.4.5, K-NN, Neural Network uses the Weka framework v.3.8.2. The results of the Decision Tree C.4.5 algorithm have a higher level of accuracy with 89.66 percent accuracy.
Keywords—Web Phising; Classification; Data Mining.

PENDAHULUAN

Berdasarkan laporan Wearesosial memuat beberapa fakta yang diantaranya tentang jumlah pengguna internet dunia telah mencapai 4,021 miliar orang. Di Indonesia sendiri, dijelaskan bahwa pengguna internet ditanah air mencapai 132 juta orang jumlah ini menunjukan lebih dari 50% penduduk Indonesia telah bisa mengakses internet. Sementara di laporan yang sama dijelaskan dari ratusan juta pengguna internet di Indonesia tersebut 60% persennya telah mengakses internet menggunakan ponsel pintar (smartphone). Tidak hanya dari segi akses yang terus meningkat, tetapi juga dari durasi menggunakan internet. Wearesocial melaporkan bahwa rata-rata dunia menggunakan internet selama enam jam per hari untuk mengakses internet melalui berbagai perangkat. Jika durasi ini dikalikan dengan jumlah pengguna internet dunia, maka durasi penggunaan internet oleh seluruh manusia di bumi bisa mencapai lebih dari 1 miliar jam untuk online di tahun 2018. Penggunaan internet tersebut didominasi oleh aktifitas bersosialisasi di dunia maya. Terbukti dengan jumlah pengguna sosial media dunia yang begitu besar jumlahnya, mencapai 3,196 miliar pengguna. Indonesia dalam hal jumlah pengguna sosial media mencapai 49% persen populasi pengguna internet atau hampir separuh pengguna internet di Indonesia telah memiliki sosial media[1].

Namun dengan meningkatnya pengguna internet dan sosial media disertai juga dengan maraknya web phising. Meskipun sebagian pengguna internet menyadari akan maraknya serangan web phising tersebut. Tujuan dari serangan ini adalah membuat

pengguna percaya bahwa mereka berinteraksi dengan situs resmi dunia maya atau online yang dimaksud. Umumnya informasi yang dicari phiser(pelaku phising) adalah berupa password, akun atau nomer kartu kredit korban dengan cara diarahkan ke web palsu atau mengirim email,banner atau pop-up dimana korban diminta untuk memberikan infomasi pribadinya. Oleh karna itu perlu adanya deteksi web phising yang berguna melindungi dari data sensitif pengguna.

Data mining merupakan bidang ilmu yang dapat digunakan untuk mendeteksi web phising. Data mining merupakan sebuah konsep untuk mengenali pola yang tersembunyi dan menemukan relasi antar parameter didalam data dengan jumlah yang besar[2].Didalam data mining terdapat beberapa teknik antara lain estimasi,klastering,asosiasi, prediksi, dan klasifikasi. Salah satu teknik yang memungkinkan untuk digunakan ialah klasifikasi. Klasifikasi merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Pada kasus deteksi web phising, teknik klasifikasi digunakan untuk mengklasifikasi web yang berpotensi adanya serangan web phising dan web yang aman.

Klasifikasi web tersebut berdasarkan pada beberapa ciri atau gejala seperti Using the IP Address, Long URL, URL having @ Symboly, Adding Prefix and Suffix, Sub-Domain(s),Misuse of HTTPs, Request URL, URL of Anchor, Server Form Handler, Abnormal URL, Redirect Page, Using Pop-up Window, Hiding Suspicious Link, DNS record, Website Traffic, Age of Domain, Disabling Right Click[3].

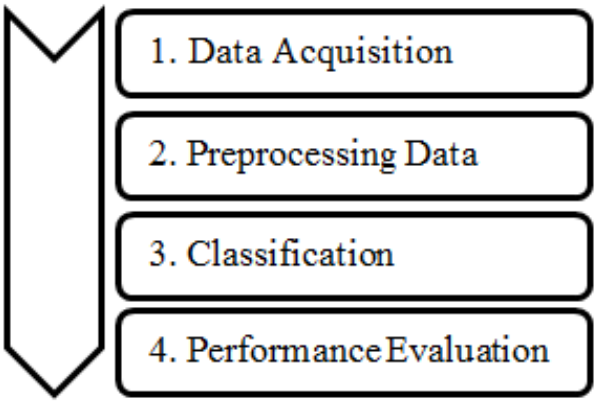
Secara umum algoritma klasifikasi yang sering digunakan ialah Decision Tree, Neural Network, k-Nearest Neighbor, C 4.5, ID3, dan Naive Bayes. Seperti penelitian yg dilakukan oleh Tony Salim dan kawan-kawan menggunakan algoritma Decision Tree(C 4.5) dalam menangani klasifikasi web phising, hasil penelitian tersebut menyatakan bahwa algoritma Decision Tree dapat menangani permasalahan klasifikasi web phising dengan baik yaitu dari 30 atribut yang digunakan yang berpengaruh dalam logika Decision Tree adalah 9 atribut yaitu SSLfinal_State, Preffix_Suffix, URL_of_Anchor,having_Sub_Domain,web_traffic,having_IP_Address,Domain_registration_length, Links_in_tags, dan iframe.

Selamet Widodo melakukan penelitian terkait deteksi web phising dengan menggunakan teknik klasifikasi. Dalam penelitian tersebut Selamet Widodo membandingkan algoritma Neural Network dan K-Nerest Neighbor dimana hasil menunjukkan bahwa algoritma Neural Network memberikan hasil klasifikasi terbaik dengan menggunakan nilai parameter training cycle=1000, learning rate=0.3 dan momentum=0.3 dengan ketepatan klasifikasi sebesar 91.21% dengan AUC 0.969. Sedangkan metode KNN memberikan hasil performansi klasifikasi terbaik saat k = 6 dengan hasil ketepatan klasifikasi 90.33%. Neural network dengan backpropagation mampu memberikan hasil klasifikasi yang lebih baik dibandingkan dengan metode KNN.Memiliki kinerja yang lebih baik dalam menangani deteksi web phising[4].

Berdasarkan uraian tersebut, pada penelitian ini akan dilakukan perbandingan algoritma C.4.5, Neural Network dan K-Nerest Neighbor untuk mengetahui kinerja dari algoritma tersebut dalam mendeteksi webphising dengan menggunakan teknik klasifikasi. Sehingga akan diperoleh satu algoritma klasifikasi yang memiliki kinerja paling baik untuk menangani kasus ini. Dengan demikian, algoritma klasifikasi yang memiliki kinerja paling baik tersebut nantinya dapat diimplementasikan ke dalam sebuah sistem pendeteksi webphising yang secara otomatis memiliki kemampuan yang baik karena di dalamnya terdapat algoritma yang memiliki kinerja yang baik.

METODE

Penelitian klasifikasi *webphising* ini dilakukan berdasarkan tahapan penelitian yang meliputi *Data Acquisition, Data Preprocessing, Classification Process,* serta *Evaluation* sebagaimana ditunjukkan pada gambar 1.



Gambar 1. Diagram Alur Tahapan Klasifikasi Web Phising

Berikut ini adalah pemaparan dari langkah-langkah penelitian yang ditunjukkan pada Gambar 1.

Data Acquisition

Dataset yang digunakan pada penelitian ini adalah data Website Phising Data Set.https://archive.ics.uci.edu/ml/datasets/Website+Phishing. Dataset tersebut merupakan dataset publik yang diambil dari UCI Machine Learning. Data Website Phising Data Setmemiliki X instance, 10 attributes fitur dan 1 attribute class. Berikut Tabel 1. Merupakan detail attribute Website Phising Data Set yang digunakan pada penelitian ini.

TABEL 1 FEARTURE DATASET

Feature	Value
Server Form Handler	1,0,-1
Using Pop-up Window	
SSL Final State	
Request URL	
Website Traffic	
URL Length	
Age of Domain	1,-1
Having IP Address	0,1
Result	1,0,-1

Berikui ini merupakan penjelasan dari fitur pada dataset [4]:

1) Server Form Handler

SFH yang mengandung string kosong atau "about: blank" dianggap meragukan karena akan mengambil informasi yang disampaikan. Selain itu, jika nama domain di SFH berbeda dari nama domain dari halaman web, ini mengungkapkan bahwa halaman web adalah curiga karena informasi yang disampaikan jarang ditangani oleh domain eksternal.

2) Using Pop-up Window

Hal ini yang tidak wajar jika situs yang sah meminta pengguna untuk mengirimkan identitasnya melalui jendela popup. Jika hal ini terjadi maka diklasifikasikan ke "phishing".

3) SSL Final State

SSL atau Secure Socket Layer adalah cara sebuah situs web membuat sambungan aman dengan browser web pengguna. Setiap kali seorang surfer web mengunjungi situs yang aman yang menggunakan teknologi SSL, menciptakan sebuah link yang terenkripsi antara sesi browser mereka dan web server. SSL adalah standar industri untuk komunikasi web yang aman dan digunakan untuk melindungi jutaan transaksi online setiap hari.

4) Request URL

Sebuah halaman web biasanya terdiri dari teks dan beberapa objek seperti gambar dan video. Biasanya, objek-objek ini dimuat ke halaman web dari domain yang sama di mana halaman web yang ada. Jika objek yang diambil dari domain yang berbeda dari domain yang diletik di alamat URL maka halaman web adalah berpotensi mencurigakan.

5) Website Traffic

Fitur ini mengukur popularitas website dengan menentukan jumlah pengunjung dan jumlah halaman yang mereka kunjungi. Situs phising tidak ditemukan dalam database Alexa. Situs yang sah memiliki rentang 100.000 peringkat. Selain itu, jika domain tidak memiliki lalu lintas atau tidak diakui oleh database Alexa, diklasifikasikan sebagai "Phishing". Jika bukan keduanya diklasifikasikan sebagai "Mencurigakan".

6) URL Length

Serangkaian karakter (dapat berupa angka, huruf, atau pun simbol) yang disusun sedemikian rupa dengan mengikuti standar yang sudah ditentukan sebelumnya dan berfungsi sebagai penunjuk alamat sumber daya dokumen yang ada di internet.

7) Age of Domain

Situs web dianggap "sah" jika usia domain lebih dari 2 tahun.

8) Having IP Address

IP Address adalah alamat atau identitas numerik yang diberikan kepada sebuah perangkat komputer agar komputer tersebut dapat berkomunikasi dengan komputer lain.

9) Result

Pengklasifikasian web apakah dia masuk aman, mencurigakan maupun terinfeksi phising.

Preprocessing Data

Data preprocessing merupakan tahapan dimana data akan dilakukan pengisian data yang kosong, menghilangkan duplikasi data, memeriksa inkonsistensi data, pembersihan data serta memperbaiki kesalahan pada data. Proses pembersihan meliputi pengisian data yang kosong, menghilangkan duplikasi data, memeriksa inkonsistensi data, dan memperbaiki kesalahan pada data. Biasanya data yang kosong disebabkan oleh adanya data baru yang belum ada informasinya [5].

Pada penelitian ini dataset yang digunakan kebetulan tidak terdapat missing value atau data yang kosong. Sehingga bisa langsung dilanjutkan ketahapan berikutnya.

Classification

Model klasifikasi deteksi situs phising pada penelitian ini mengacu pada penelitian yang dilakukan oleh Selamet Widodo. Akan tetapi pada tahap pada penelitian kami dilakukan perbandingan algoritma sebelumnya kami bandingkan siapa yang paling terbaik. Pada tahap ini akan difokuskan untuk membandingkan empat algoritma klasifikasi yaitu Naïve Bayes,C.4.5,K-NN,Neural Network..

1) K-Nearest Neighbor

K-Nearest Neighbor merupakan sebuah algoritma klasifikasi yang bekerja berdasarkan jarak terpendek dari query instance ke *trainingsample* untuk menentukan KNN-nya. *Training* sample diproyeksikan ke dalam ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang tersebut dibagi menjadi beberapa bagian berdasarkan klasifikasi *trainingsample*. Sebuah titik yang terdapat pada ruang ini akan ditandai sebagai kelas c apabila kelas c merupakan klasifikasi yang paling banyak ditemui pada *k* tetangga terdekat dari titik tersebut. Jarak antara tetangga biasanya dihitung berdasarkan *Euclidean Distance* yang direpresentasikan sebagai berikut [6] :

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2} \qquad (1)$$

dimana,

$d(x_i, x_j)$: Jarak *Euclidean* (*Euclidean Distance*)

(x_i) : *record* ke- i

(x_j) : *record* ke- j

(a_r) : data ke-r

i, j : 1,2,3,…n

2) Neural Network

Neural Network banyak diterapkan dalam penelitian karena kemampuannya dalam memodelkan sistem yang sangat nonlinier di mana hubungan antara variabel-variabel tidak diketahui (generalisasi) atau sangat kompleks[7]. Kemampuan dari ANN telah dibuktikan pada beberapa aplikasi termasuk speech synthesis, diagnosa, bidang pengobatan, keuangan dan bisnis, kontrol pada robot, pemrosesan sinyal, dan masalah lain yang termasuk dalam kategori pengenalan pola dan klasifikasi[8]. Teknik yang populer digunakan pada metode ANN adalah algoritma Back Propagation (BP)

3) Naive Bayes

Naive Bayes merupakansebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi tidak ada ketergantungan (*independent*) yang tinggi. Salah satu keuntungan algoritma Naive Bayes ialahdalam menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian hanya membutuhkan jumlah data pelatihan yang kecil. Karena diasumsikan sebagai variable independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians[6]. Berikut ini persamaan umum dari Naive Bayes :

$$p(C|F_1,\dots,F_1) = \frac{p(C)\,p(F_1,\dots,F_n|C)}{p(F_1,\dots,F_n)} \quad (2)$$

Dimana variabel C merepresentasikan kelas, sementara variabel F_1...Fn merepresentasikan berbagai karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka persamaan tersebut menjelaskan bahwa peluang masuknya sampel dengan karakteristik tertentu dalam kelas C adalah peluang munculnya kelas C sebelum masuknya sampel tersebut, dikali dengan peluang kemunculan berbagai karakteristik sampel pada kelas C dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (*evidence*).

4) Decision Tree C4.5

Algoritma C4.5 disebut juga decision tree atau pohon keputusan karena algoritma ini menghasilkan pohon keputusan. Algoritma C4.5 merupakan algoritma berbasis rule yang diperoleh secara tidak langsung, karena rule diperoleh dari pohon keputusan yang dihasilkan oleh algoritma C4.5.[9]. Algoritma C4.5 merupakan pengembangan dari ID3 yang dikembangkan oleh J. R. Quintlant pada tahun 1987 [6]. Untuk membangun pohon keputusan dalam algoritma C4.5, hal pertama yang dilakukan yaitu memilih atribut sebagai akar, kemudian dibuat cabang untuk tiap-tiap nilai didalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Untuk memilih atribut dengan akar, didasarkan pada nilai gain tertinggi dari atribut-tribut yang ada. Gain (S,A) merupakan perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variable dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A). Berikut ini persamaan untuk menghitung nilai gain [10].

$$Gain\ (S,A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\ (Si) \quad (3)$$

dimana,

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Sedangkan nilai Entropy dapat dihitung menggunakan persamaan berikut ini.

$$Entropy\ (S) = \sum_{i=1}^{n} - pi * log_2 pi \quad (4)$$

dimana,

S : Himpunan kasus

n : Jumlah partisi S

pi : Proporsi dari Si terhadap S

Performance Evaluation

Evaluasi merupakan proses pengujian kinerja algoritma klasifikasi yang digunakan. Pada umumnya evaluasi kinerja algoritma klasifikasi. menggunakan confusion matrix[11]. Evaluasi dengan confusion matrix akan menghasilkan nilai accuracy, sensitivity dan specificity. Akurasi dalam klasifikasi merupakan persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [14]. Specificity proporsi kasus negatif yang diidentifikasi dengan benar. Recall atau sensitivity merupakan proporsi kasus positif yang diidentifikasi dengan benar[12]. Sedangkan F-Measure merupakan salah satu perhitungan evaluasi yang mengkombinasikan recall dan precission,[13]. Berikut ini Tabel 2 merupakan tabel confusion matrix.

TABEL 2 CONFUSION MATRIX [6]

Actual Classification	*Prediction Classificatiom*	
	Positif	Negatif
Positif	True Positif	False Negatif
Negatif	False Positif	True Negatif

Berdasarkan Tabel 2, *True* Positive adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false* positif adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false* negatif adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true* negatif adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung nilai*sensitivity* (*recall*), *precision*, *f-measure* dan*accuracy*. Untuk menghitung nilai-nilai tersebut digunakan persamaan dibawah ini [14].

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

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

$$Sensitivity\ (recall) = \frac{TP}{(TP+FN)} \quad (5)$$

$$F - Measure = 2\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (6)$$

HASIL DAN PEMBAHASAN

Pada penelitian ini, kami akan melakukan perbandingan performa dari lima algoritma klasifikasi, Kelima algoritma tersebut ialah Naïve Bayes,C 4.5,K-NN,Neural Network,dan Suport Vector Machine. Alat bantu yang digunakan ialah WEKA v.3.8.2 Untuk validasi digunakan metode 80% sebagai data training dan sisanya digunakan sebagai data set.

Berikut ini table III, IV, V, VI merupakan table confusion matrix hasil dari pengolahan data menggukan algoritma k-NN, Naïve Bayes,C4.5 dan Neural Network dengan 0 sebagai kelas mencurigakan, 1 sebagai kelas valid (atau web yang bersih dari phising), -1 sebagai kelas terdeteksi terkena web phising.

TABEL 3 CONFUSION MATRIX K-NEAREST NEIGHBOR

Actual Classification	*Prediction Classificatiom*		
	0	1	-1
0	11	1	2
1	4	98	8
-1	3	17	127

TABEL 4 CONFUSION MATRIX NEURAL NETWORK

Actual Classification	*Prediction Classificatiom*		
	0	1	-1
0	10	1	3
1	2	99	9
-1	2	13	132

TABEL 5 CONFUSION MATRIX NAIVE BAYES

Actual Classification	*Prediction Classificatiom*		
	0	1	-1
0	0	3	11
1	6	97	7
-1	1	17	129

TABEL 6 CONFUSION MATRIX C4.5

Actual Classification	*Prediction Classificatiom*		
	0	1	-1
0	12	0	2
1	3	98	9
-1	2	12	133

Dengan menggunakan tabel confusion matrix algoritma k-NN, Naïve Bayes, C4.5 dan Neural Network diatas maka dapat dihitung nilai akurasi, presisi, *sensitivity* dan *f-measure* dengan menggunakan persamaan (3), (4) (5) dan (6).

Tabel VII berikut ini menunjukan perbandingan tingkat akurasi, sensitifity, presisi, dan f-measure hasil klasifikasi dari kelima algoritma yang digunakan. Algoritma C 4.5 memiliki performa yang paling baik dari sisi akurasi, sensitifity,presisi, dan f-Measure.

TABEL 7 PERBANDINGAN KINERJA

Algoritma	Akurasi	Sensitifity	Presisi	F-Measure
Naive Bayes	83,39	0,834	0,834	0,823
C 4.5	89,66	0,897	0,899	0,897
KNN	87,08	0,871	0,877	0,873
Neural Network	88,92	0,889	0,890	0,889

Pada table diatas dapat dilihat nilai ketepatan klasifikasi paling optimal adalah algoritma C 4.5 dengan Akurasi sebesar 89,66% ,Sensitifity 0,897% , Presisi 0,899% , dan F-Measure 0,897%. Yang artinya algoritma Decision Tree C 4.5 baik digunakan dalam mengklasifikasikan situs web phising.

KESIMPULAN

Berdasarakan pengujian model berbasiskan metode Naïve Bayes, Decision Tree C4.5, K-NN, Neural Network menggunakan framework Weka v.3.8.2. Didapat hasil penelitian dari 1353 dataset yang diolah dan menggunakan 10 atribut membuktikan bahwa algoritma Decision Tree C4.5 memiliki kinerja yang lebih tinggi dibandingkan dengan tiga algoritma lainnya yaitu Neural Network, Naïve Bayes dan k-NN. Hal ini karena secara keseluruhan algoritma C4.5 mempunyai nilai kinerja baik akurasi, sensitifity, presisi maupun f-measure yang paling tinggi.

SARAN

Untuk penelitian selanjutnya dengan permasalahan yang sama, akan dilakukan penanganan terhadap imbalance class pada dataset yang mungkin dapat memberikan hasil yang lebih baik..

DAFTAR PUSTAKA

[1] "Digital in 2018: World's internet users pass the 4 billion mark - We Are Social," *wearesosial.com*. [Online]. Available: https://wearesocial.com/blog/2018/01/global-digital-report-2018. [Accessed: 20-Nov-2018].

[2] P. Kaur, M. Singh, and G. S. Josan, 2015, "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector," in *Procedia Computer Science*.

[3] R. M. Mohammad, F. Thabtah, and L. McCluskey, 2014, "Predicting phishing websites based on self-structuring neural network," *Neural Comput. Appl.*

[4] S. Widodo, 2017, "Klasifikasi Situs Phishing dengan Menggunakan Neural Network dan K-Nearest Neighbor," *Inf. Manag. Educ. Prof.*, vol. 1, no. 2, pp. 145–154.

[5] O. N. Pratiwi, 2013, "Predicting student placement class using data mining," in *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2013*.

[6] Jiawei Han and Micheline Kamber, 2006, *Jiawei Han & Micheline Kamber*, Second Edi. San Francisco: Morgan Kaufmann Publishers.

[7] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, 2013, "Artificial neural networks in medical diagnosis," *Journal of Applied Biomedicine*.

[8] M. P. Kinjal Jadav, "OPTIMIZING WEIGHTS OF ARTIFICIAL NEURAL NETWORKS USING GENETIC ALGORITHMS."

[9] X. Yu, L., Chen, G., Koronios, A., Zu, S., & Guo, 2007, "Application and Comparison of Classification Techniques in Controlling Credit Risk," *Recent Adv. Data Min. Enterp. Data*, vol. 111–146.

[10] E. T. L. Kusrini, 2009, *Algoritma Data Mining*.

[11] M. Bramer, 2007, *Principles of Data Mining, Undergraduate Topics in Computer Science*.

[12] D. M. W. Powers, 2011, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*

[13] Y. Pristyanto, I. Pratama, and A. F. Nugraha, 2018, "Data level approach for imbalanced class handling on educational data mining multiclass classification," in *2018 International Conference on Information and Communications Technology, ICOIACT 2018*.

[14] M. Han, J., & Kamber, 2006, *Data Mining: Concepts and Techniques Second*, Second Edi., vol. 12. San Fransisco: Morgan Kauffman.