Pengaruh Jumlah Hiden Layer Terhadap Performa Neural Network Dengan Dalam Prediksi Website Phishing

Allan D Alexander

Teknik Informatika, Universitas Bhayangkara Jakarta Raya

allan@ubharajaya.ac.id

Abstrak

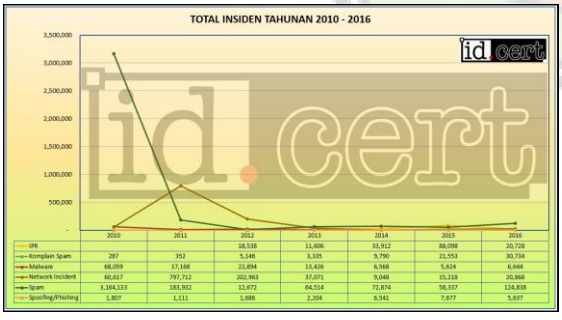
Berdasarkan laporan dari ID-CERT jumlah insiden phishing dari tahun 2010 sampai dengan tahun 2016 cenderung meningkat, insiden yang tercatat pada tahun 2010 yaitu 1807 insiden dan 5607 pada tahun 2016. Banyak usaha yang dilakukan untuk memprediksi website phishing dengan menngunakan teknik data mining. *Neural network* sudah menunjukan performanya sebagai penaksir yang sangat umum, optimasi jumlah *neuron* yang tersembunyi tanpa menentukan akurasi sebelumnya merupakan tantangan utama bagi penggunaan *neural network*

Keywords

Neural network, website phising, confusion matrix, data mining

Pendahuluan

Berdasarkan laporan dari ID-CERT jumlah insiden *phishing* dari tahun 2010 sampai dengan tahun 2016 cenderung meningkat, insiden yang tercatat pada tahun 2010 yaitu 1807 insiden dan 5607 pada tahun 2016 [1].



Gambar 1Total Insiden Tahun 2010 – 2016
Sumber: ID-CERT

Phishing merupakan tindakan penipuan dimana pengguna email akan diarahkan pada sebuah website untuk memberikan informasi pribadi atau informasi yang bersifat rahasia kepada pihak-pihak tertentu dengan tujuan pelanggaran hukum (contoh: pencurian identitas) [2].

Upaya untuk memprediksi website phishing merupakan hal yang mendasar dan dapat dilakukan dengan menggunakan neural network. Banyak usaha yang dilakukan untuk memprediksi website phishing dengan menngunakan teknik data mining, namun tingkat rata-rata kesalahan dari algoritma yang digunakan masih terlalu tinggi [3]. Neural network memiliki akurasi hingga 98,72% dalam memprediksi website phishing [4] .

Neural Network

Neural Network sesuai dengan namanya mengindikasikan, jaringan komputasi yang mencoba mensimulasikan secara umum jaringan sel saraf (*neuron*) dari sistem saraf pusat mahluk hidup (manusia atau hewan). Istilah ini dipinjam dari pengetahuan tentang neurofisiologi dibidang

biologi *neuron* dan jaringan biologi *neuron*. Berbeda dengan mesin komputasi konvensional (digital atau analog) yang berfungsi sebagai pengganti untuk meningkatkan kemampuan atau kecepatan pehitungan yang dilakukan oleh otak manusia tanpa memperhatikan organisasi dari elemen komputasi dan jaringannya. [5]

Neural network sudah menunjukan performanya sebagai penaksir yang sangat umum, optimasi jumlah *neuron* yang tersembunyi tanpa menentukan akurasi sebelumnya merupakan tantangan utama bagi penggunaan *neural network*. Penggunaan *neuron* tersembunyi yang berlebihan dapat menyebabkan ketidak sesuaian , artinya *neural network* dapat memerkirakan secara berlebihan kompleksitas dari sebuah permasalahan yang sedang ditanganinya [6].

Objek Penulisan

Paper ini akan menentukan akurasi *neural network* dalam melakukan klasifikasi website phising berdasarkan jumlah hiden layernya. Untuk mengukur akurasi yang dimaksud paper ini menggunakan dataset yang diperoleh dari UCI machine learning repository yang berisi data-data yang berkaitan dengan website phising, dengan keterangan sebagai berikut;

Tabel 1 Properti dataset

Jumlah Instance	11055
Jumlah Attribute	31

Dan untuk melakukan evaluari kinerja dari model *neural network* yang akan dibentuk akan menggunakan model *confusion matrix* dan *ROC Curve (Reiver Operating Characteristic).* Dalam paper ini semua percobaan dilakukan dengan software WEKA.

Model neural network

Paper ini membandingkan 2 buah model *neural network* dengan jumlah hiden layer yang berbeda, dimana model *neural network* pertama akan menggunakan 1 hiden layer dan model *neural network* berikutnya akan menggunakan lebih dari 1 hiden layer, dan untuk proses training dan testing data set akan dibagi menjadi dua bagian dimana 80% bagian data set akan digunakan sebagai training dan 20% lainya akan digunakan untuk testing.

Model *neural network* yang pertama memiliki properti sebagai beikut;

Tabel 2 Properti neural network model pertama

Properti	Jumlah
Jumlah input	31
Hiden Layer	1
Sigmoid	25
Learning Rate	0,3
Momentum	0,2
Training time	500
Training dataset	80%
Testing dataset	20%

Dari model *neural network* tersebut, didapatkan hasil sebagai beikut;

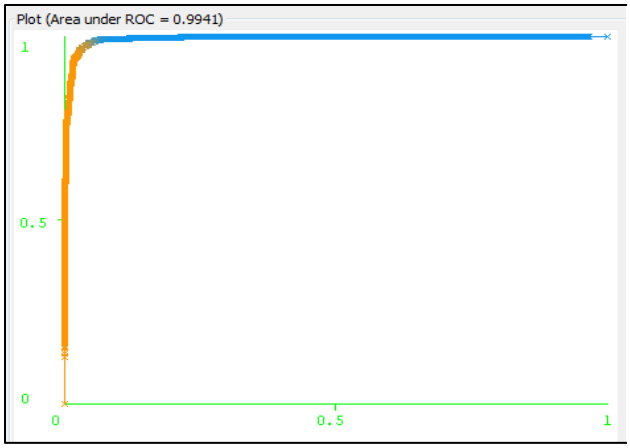
Tabel 3 hasil neural network model pertama

Deskripsi	
Waktu pembentukan model	261.46 detik
Akurasi prediksi	96.9 %

Sedangkan performa dari *neural network* model pertama dapat dilihat dari confusion matrix dan ROC-curve berikut;

Tabel 4 Confusion matrix neural network model pertama

Classification		**Predicted Class**	
		a	**b**
Observed Class	**a = -1**	*955*	*46*
	b = 1	*23*	*1187*



Gambar 2 ROC-Curve neural nerwork model pertama

Untuk neural network model kedua memiliki properti sebagai berikut;

Tabel 5 Properti neural network model pertama

Properti	Jumlah
Jumlah input	31
Hiden Layer	2
Sigmoid layer 1	49
Sigmoid layer 2	25
Learning Rate	0,3
Momentum	0,2
Training time	500
Training dataset	80%
Testing dataset	20%

Dari model *neural network* tersebut, didapatkan hasil sebagai beikut;

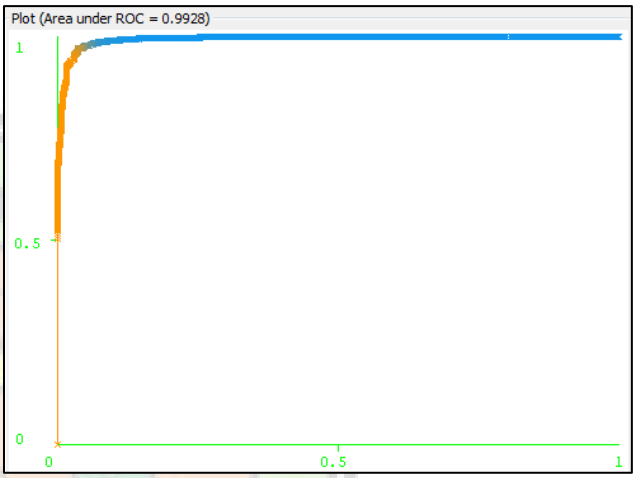
Tabel 6 hasil neural network model kedua

Deskripsi	
Waktu pembentukan model	318.32 detik
Akurasi prediksi	93.5 %

Sedangkan performa dari *neural network* model kedua dapat dilihat dari confusion matrix dan ROC-curve berikut;

Tabel 7 Confusion matrix neural network model pertama

Classification		**Predicted Class**	
		a	**b**
Observed Class	**a = -1**	*952*	*49*
	b = 1	*29*	*1181*



Gambar 3 ROC-Curve neural nerwork model kedua

Kesimpulan

Dari dua model yang telah dibentuk terlihat bahwa *neural network* model pertama yang terdiri dari satu hiden layer memiliki performa yang lebih baik, hal tersebut bisa dilihat dari nilai akurasi dalam mendeteksi website phising sebesar 96,9% dibandingkan dengan *neural network* model kedua yang terdiri dari 2 hiden layer yang memiliki nilai akurasi 93.5%. Dan untuk selanjutnya perlu penelitian lebih lanjut untuk menentukan struktur model *neural network* yang optimal untuk mendapatkan performa yang lebih baik.

Referensi

[1] Id-CERT, “ID-CERT,” IC-CERT, 2017. [Online]. Available: http://www.cert.or.id/tentang-kami/en/. [Diakses 17 March 2017].

[2] Merriam-Webster, "Merriam-Webster," 19 Maret 2017. [Online]. Available: https://www.merriam-webster.com/dictionary/phishing. [Diakses 17 Maret 2017].

[3] C. J. Chandan, H. P. Chheda, D. M. Gosar dan H. R. Shah, "A Machine Learning Approach for Detection of Phished Website Using Neural Network," *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 2, no. 2, pp. 4205-4209, 2014.

[4] N. G. Jameel dan L. E. George, "Detection of Phishing Email using Feed Forward Neural Network," *International Journal of COmputer Application,* pp. 10-15, 2013.

[5] D. Graupe, Principles Of Artificial Neural Networks 2nd Edition, Singapore: World Scientific Publishing, 2007.

[6] Y. Liu, J. A. Starzyk dan Z. Zhu, "OPTIMIZING NUMBER OF HIDDEN NEURAL NETWORKS," *AIA,* 2007.

[7] J. Jiang, "BP Neural Network Algorithm Optimized by Genetic Algorithm and Its Simulation," *International Journal of Computer Science Issue Vol. 10, Issue 1, No. 2,* pp. 516-518, 2013.

[8] R. L. Haupt dan S. E. Haupt, Practical Genetic Algorithms, New Jersey: John Wiley & Sons, 2004.