

## Klasifikasi Menggunakan Metode *Hybrid Bayesian-Neural Network* (Studi Kasus: Identifikasi Virus Komputer)

Dian C. Rini<sup>1</sup>, Yuniar Farida<sup>2</sup>, Dwi Puspitasari<sup>3</sup>

UIN Sunan Ampel, Surabaya<sup>1,2</sup>, Politeknik Negeri Malang<sup>3</sup>  
diancrini@uinsby.ac.id<sup>1</sup>, greatyuniar@gmail.com<sup>2</sup>, dwi\_sti@yahoo.com<sup>3</sup>

### Abstrak

Virus komputer merupakan suatu program yang menginfeksi komputer terutama pada saat komputer sedang beroperasi dan menjadi momok bagi pengguna komputer. Virus komputer dapat menggandakan dirinya sendiri dan menyebar dengan cara menyisipkan dirinya pada program dan data lainnya. Efek negatif virus komputer adalah memperbanyak dirinya sendiri, yang membuat sumber daya pada komputer terutama penggunaan memori menjadi berkurang secara signifikan. Diperlukan suatu penangkal atau antivirus dalam mencegah penyebaran yang lebih jauh dalam sistem komputer. Pada penelitian ini, dilakukan suatu identifikasi virus dengan menggabungkan dua metode yaitu *Naïve Bayes Classifier* dengan *Neural Network*.

Fitur virus didapatkan dari mengkodekan ciri-ciri dari virus. Untuk klasifikasi awal digunakan metode *Naïve Bayes Classifier* untuk membagi dua jenis fitur, yaitu virus dan bukan virus. Setelah masuk kedalam jenis virus, maka diklasifikasikan kedalam dua jenis virus yaitu trojan atau worm menggunakan salah satu metode *neural network (perceptron)*. Hasil sistem setelah dilakukan uji coba didapatkan *recognition rate* tertinggi yaitu sebesar 94.12%.

*Kata kunci: Virus Komputer, Naïve Bayes Classifier, Neural Network, Perceptron*

### Abstract

A computer virus is a program that infects computer, especially when the computer is running. A computer virus can reproduce itself and it spreads by inserting in the program and other data. The negative effect of a computer virus is they are multiplied, which makes computer resources, especially memory usage to be reduced significantly. In order to prevent further spread of the computer system an antidote or antiviral is needed. This study carries out an identification of the virus by combining the two methods, *Naïve Bayes Classifier* and *Neural Network*. Features of the virus is obtained by encoding the characteristics of the virus. For the initial classification Bayesian method is used to divide the two types of features, namely the virus and not the virus. The virus is classified into two types, Trojan or worm, by one of neural network (*perceptron*) methods. The results after testing the system gain recognition rate of 94.12%.

*Keywords: Virus Computer, Naïve Bayes Classifier, Neural Network, Perceptron*

## 1. Pendahuluan

Virus komputer umumnya dapat merusak perangkat lunak komputer dan tidak dapat secara langsung merusak perangkat keras komputer tetapi dapat mengakibatkan kerusakan dengan cara memuat program yang memaksa *over process* ke perangkat tertentu. Efek negatif virus komputer adalah memperbanyak dirinya sendiri, yang membuat sumber daya pada komputer (seperti penggunaan memori) menjadi berkurang secara signifikan. Hampir 95% virus komputer berbasis sistem operasi Windows. Sisanya menyerang Linux/GNU, Mac, FreeBSD, OS/2 IBM, dan Sun Operating System. Virus yang ganas akan merusak perangkat keras.[1]. Berbagai macam-macam jenis virus antara lain worm, trojan, spyware dan lain sebagainya. Ditinjau dari segi kemampuan[2], kriteria virus dibagi menjadi 5 hal antara lain:

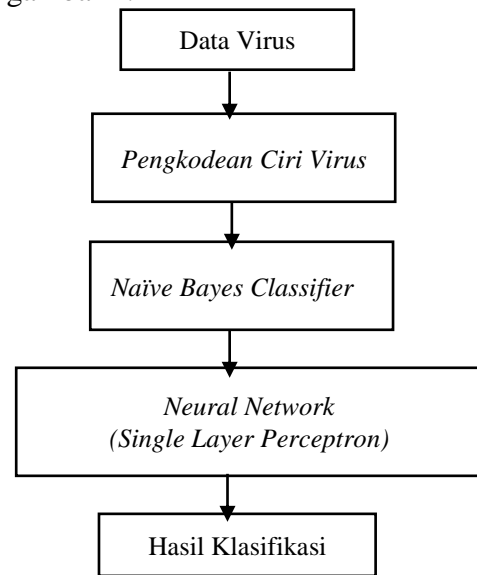
1. Kemampuan untuk mendapatkan informasi
2. Kemampuan memeriksa suatu program
3. Kemampuan untuk menggandakan diri
4. Kemampuan mengadakan manipulasi
5. Kemampuan Menyembunyikan diri

Penggolongan atau pengelompokan terhadap virus dapat dilakukan dengan beberapa metode klasifikasi. Beberapa metode yang sangat bagus menggunakan naïve bayes antara lain digunakan dalam klasifikasi deteksi email spam[3], klasifikasi Sinyal EEG[4], klasifikasi kanker payudara [5] dan *pattern recognition* yang lainnya [6]. Naïve bayes sangat cocok dan baik dalam *binary classification*. Terdapat pula banyak metode klasifikasi, antara lain *single layer perceptron*, *multi layer perceptron* dan *backpropagation*. Bahkan perceptron masih banyak digunakan dalam beberapa penelitian, antara lain klasifikasi sinyal otak (EEG)[7], handal juga dalam prediksi[8][9][10] dan *pattern recognition*[11][12]. Perceptron sangat bagus dalam *binary classification* sehingga cocok apabila dilakukan klasifikasi kedalam dua kelas, Pada penelitian ini, akan dilakukan suatu identifikasi virus berdasarkan kriteria yang ada. Naïve Bayes digunakan untuk mengklasifikasikan apakah fitur yang

diinputkan masuk kedalam kelas virus atau bukan virus, kemudian jika masuk virus maka akan dilakukan klasifikasi menggunakan metode *single layer perceptron* kedalam dua kelas yaitu virus jenis worm atau jenis trojan.

## 2. Metodologi

Metode yang diterapkan dalam penelitian ini, dapat dilihat pada gambar flowchart pada gambar 1.



Gambar 1. Diagram alur sistem proses klasifikasi virus

### 2.1 Ekstraksi Fitur

Didapatkan data berdasarkan kriteria virus, yaitu mendapatkan informasi, memeriksa program, menggandakan diri, memanipulasi dan menyembunyikan diri (*hidden*) sesuai dengan gambar 2.

Informasi	Program	MenggandakanDiri	Memmanipulasi	Hidden	Virus
Mendapatkan	Memeriksa	Tidak	Beberapa	Hidden	Ya
Mendapatkan	Memeriksa	Ganda	Tidak	Muncul	Tidak
Mendapatkan	Memeriksa	Ganda	Beberapa	Muncul	Ya
Mendapatkan	Memeriksa	Ganda	Tidak	Hidden	Ya
Mendapatkan	Tidak	Tidak	Beberapa	Hidden	Ya
Diam	Tidak	Tidak	Tidak	Muncul	Tidak

Gambar 2. Data Awal Virus

Dari data yang ada perlu dikodekan untuk mempermudah dalam proses klasifikasiya. Proses pengkodean dilakukan dengan menggunakan kode biner yaitu 0 dan 1. Untuk parameter yang pertama yaitu mendapatkan informasi, jika mendapatkan maka dikodekan 1 dan jika diam dikodekan

0. Pada parameter kedua yaitu Program, jika bersifat memeriksa program maka dikodekan 1 dan jika tidak memeriksa program maka dikodekan 0. Parameter selanjutnya yaitu menggandakan diri, jika tidak ganda maka dikodekan 0 dan jika bersifat menggandakan diri maka dikodekan 1. Pada parameter Manipulasi apabila bersifat memanipulasi maka dikodekan 1 dan jika tidak dikodekan 0, begitu juga dengan parameter *Hidden*, jika dapat melakukan menyembunyikan file maka dikodekan dengan 1 dan jika tidak dikodekan 0. Berlaku pula dengan parameter yang lain, sehingga didapatkan hasil pengkodean pada tabel 1.

Tabel 1. Data Hasil Pengkodean

Infor- masi	Pro- gram	Ganda Diri	Mani- pulasi	Hidden	Virus
1	1	0	1	1	1
1	1	1	0	0	0
1	1	1	1	0	1
1	1	1	0	1	1
1	0	0	1	1	1
0	0	0	0	0	0

## 2.2 Klasifikasi

Untuk proses klasifikasi terdapat dua metode yaitu *Naïve Bayes* dan *Single Layer Perceptron*.

### 2.2.1 Naïve Bayes Classifier

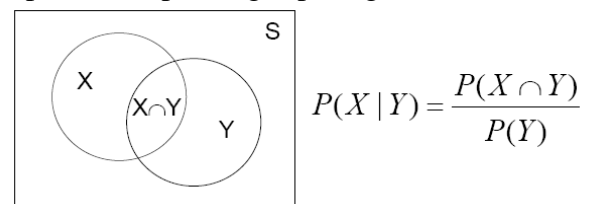
Dalam mengklasifikasikan vector fitur kedalam kelas virus atau bukan virus ini dengan menggunakan teknik *Naïve Bayes Classifier* untuk menentukan peluang kemungkinan besar termasuk kedalam kelas virus atau bukan virus.

Metode *Naïve Bayes* merupakan metode yang digunakan memprediksi probabilitas. Sedangkan klasifikasi Bayes adalah klasifikasi statistik yang dapat memprediksi kelas dari suatu anggota probabilitas. Untuk klasifikasi Bayes sederhana yang lebih dikenal sebagai *Naïve Bayesian Classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. *Naïve Bayes Classifier* merupakan sebuah metoda

klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi / kejadian, dimana diasumsikan bahwa setiap atribut contoh (data sampel) bersifat saling lepas satu sama lain berdasarkan atribut kelas. *Naïve Bayes Categorical* adalah *Naïve Bayes* dengan data statik berupa kategori atau merupakan data pasti, sehingga dalam pengerjaannya sudah didapatkan hasil yang pasti. *Naïve bayes* merupakan metode dengan rumus dasar bayesian, Pada teorema Bayes, bila terdapat dua kejadian yang terpisah (misalkan A dan B), maka teorema Bayes sebagai berikut[3]:

$$P(A|B) = (P(B|A) * P(A))/P(B) \quad (1)$$

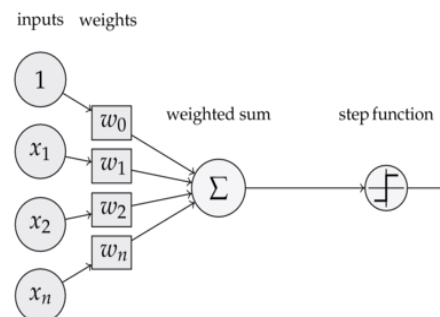
Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Representasi peluang seperti gambar 3.



Gambar 3. Diagram Venn Probabilitas

### 2.2.2 Single Layer Perceptron

*Single Layer Perceptron* merupakan salah satu metode klasifikasi *Neural Network*, dimana merupakan metode yang meniru sistem kerja dari jaringan syaraf atau yang biasa disebut Jaringan Syarat Tiruan (JST).



Gambar 4. Arsitektur *Single Layer Perceptron*  
*Perceptron* sendiri biasanya memiliki layer yaitu layer input dan output. Layer input

menunjukkan informasi dasar yang kemudian digunakan dalam Jaringan Syaraf Tiruan (*Neural Network*). Aktifitas setiap unit-unit lapisan tersembunyi (*hidden layer*) ditentukan oleh aktifitas dari unit-unit input dan bobot dari koneksi antara unit-unit input dan unit-unit lapisan tersembunyi. Karakteristik dari unit-unit output tergantung dari aktifitas unit-unit layer tersembunyi dan bobot antara unit-unit layer tersembunyi dan unit-unit output.

Arsitektur single layer perceptron, seperti pada gambar 4, layer pertama merupakan input dari jaringan perceptron dapat direpresentasikan dengan  $x$ , yaitu  $x_1$  sampai dengan  $x$  ke  $n$ . Kemudian diberikan bobot (*weights*) dan sebuah bias ditambahkan ke dalam jaringan dengan angka pembelajaran (*learning rate*). Output yang dihasilkan adalah  $Y$ . Karena metode ini adalah supervised learning maka ada threshold yang harus dilewati. Berikut adalah beberapa langkah-langkah keseluruhan metode *single layer perceptron*:

1. Inisialisasi input, bias dan bobot.

$$X_1 - X_n = \text{input ke } 1 \text{ sampai } n, \quad (2)$$

$b = \text{bias}$ ,

$\alpha = \text{angka pembelajaran } (0 < \alpha \leq 1)$ ,

$w_n = \text{bobot masing-masing input}$ ,

$t = \text{batas ambang / threshold}$

2. Mengitung Output, dengan rumus

$$Y_{in} = \sum w_n \cdot x_n + \alpha \cdot b \quad (3)$$

3. Output dalam bentuk biner dapat dinotasikan

(y) adalah 1 jika  $1 \leq Y_{in}$  dan

0 jika  $Y_{in} \leq 0$  (4)

4. Perubahan nilai bobot ( $w$ ) dan bias ( $b$ ) selama *learning* dinotasikan sebagai berikut:

$$W_{\text{baru}} = W_{\text{lama}} + X, \text{ jika } Y < t \quad (5)$$

$$W_{\text{baru}} = W_{\text{lama}} - X, \text{ jika } Y > t \quad (6)$$

$$W_{\text{baru}} = W_{\text{lama}}, \text{ jika } Y = t \quad (7)$$

### 3. Pembahasan Hasil

Aplikasi klasifikasi menggunakan metode *Hybrid Bayesian-Neural Network*

untuk identifikasi virus ini menggunakan *development tool* Visual Basic 6.0. Untuk klasifikasi, pertama dilakukan untuk menggolongkan data kedalam virus dan bukan virus. Apabila, ada data baru Mendapatkan informasi, Menggandakan diri dan memeriksa program, apakah virus atau bukan. Sebelumnya kita menggunakan data pada tabel 1 dan menghitung propabilitas pada masing-masing peluang kejadian menggunakan rumus 1.

Fakta:

$$P(\text{virus}=ya) = 4/6, \quad P(\text{virus}=bukan) = 2/6$$

$$P(\text{menggandakan}=ya | \text{virus}=ya) = 1/2,$$

$$P(\text{menggandakan}=ya | \text{virus}=bukan) = 1/2$$

$$P(\text{memeriksa-prog}=ya | \text{virus}=ya) = 3/4$$

$$P(\text{memeriksa-prog}=ya | \text{virus}=bukan) = 1/2$$

HMAP dari keadaan ini dapat dihitung dengan:

Probabilitas Virus = Ya,

$$P(\text{menggandakan}=ya, \text{memeriksa-prog}=ya | \text{virus}=ya) = \{P(\text{menggandakan}=ya | \text{virus}=ya) \cdot P(\text{memeriksa-prog}=ya | \text{virus}=ya)\} \cdot P(\text{virus}=ya) = \{(1/2) \cdot (3/4)\} \cdot (4/6) = 1/4$$

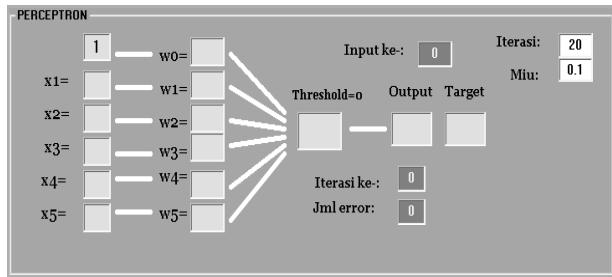
Probabilitas Virus = Tidak,

$$P(\text{menggandakan}=ya, \text{memeriksa-prog}=ya | \text{virus}=bukan) = \{P(\text{menggandakan}=ya | \text{virus}=bukan) \cdot P(\text{memeriksa-prog}=ya | \text{virus}=bukan)\} \cdot P(\text{virus}=bukan) = \{(1/2) \cdot (1/2)\} \cdot (2/6) = 1/12$$

Nilai probabilitas terhadap virus nilainya lebih besar dari probabilitas bukan virus, sehingga data yang dimasukkan tadi terdeteksi sebagai virus. Telah diuj coba terhadap 67 data. Dari 67 data, yang terklasifikasi kedalam virus sebanyak 52 dan bukan virus 25. Data yang terklasifikasi masuk ke virus, akan digolongkan menjadi dua yaitu apakah trojan atau worm menggunakan *single layer perceptron*.

Antarmuka atau tampilan dari metode *single layer perceptron* menggunakan *development tool* Visual Basic 6.0, inputan dari  $x_1$  sampai dengan  $x_5$  sesuai dengan parameter input yang ada pada tabel 1 yaitu mendapatkan informasi, memeriksa program, menggandakan diri, memanipulasi data, dan menyembunyikan diri (*hidden*). Di bangkitkan secara random pembobot antara 0-1, sesuai dengan jumlah input di inisialisasi

bobot berjumlah 5 dan ditambah satu bobot ( $w_0$ ) sebagai pengali dari bias. Kemudian dihitung outputnya dan dilakukan peng-update bobot sesuai dengan aturan pada rumus 5,6 dan 7.



Gambar 5. Antarmuka Single Layer Perceptron

Struktur jaringan yang digunakan sebagai training dan testing seperti gambar 5, telah dilakukan uji / testing dengan menggunakan parameter uji *recognition rate* berikut.

$$Recognition\ rate = \frac{Jumlah\ terklasifikasi\ benar}{Jumlah\ Data} \times 100,$$

Sebanyak 52 data, diambil secara random untuk data training dan testingnya, didapatkan seperti tabel 2 berikut.

Tabel 2. Hasil Uji Training dan Testing Program

Data Training	Data Testing	Jumlah Terklasifikasi Benar	Recognition Rate (%)
30	22	20	90,90
35	17	16	94,12
27	25	23	92,00
40	12	11	91,67
32	30	28	93,33

Dari hasil uji coba 5 kali training dengan random dan perbedaan jumlah data training dan data testing, rata-rata *recognition rate*-nya diatas 90%, bahkan tertinggi dengan data training sejumlah 40, data testing 12 dengan 11 terklasifikasi benar sehingga *recognition rate* nya tinggi yaitu sebesar 94,12%.

#### 4. Kesimpulan

Dari sistem yang telah dibuat, pada tabel 2 menyatakan bahwa rata-rata recognition ratenya diatas 90%, sehingga metode ini

sudah sangat mampu untuk mengenali pola atau dengan kata lain mampu mengidentifikasi virus atau bukan virus, serta worm atau trojan.

Dari 5 kali uji coba training dan testing, didapatkan recognition rate sistem sebesar 94,12%, hal ini berarti metode hybrid ini mampu mengidentifikasi jenis virus, hal tersebut dikarenakan data yang digunakan biner (0 dan 1) dimana kedua metode yang di hybrid sangat mampu melakukan klasifikasi yang pemisahannya secara linear.

Penelitian ini masih belum bisa dikatakan sempurna, karena masih dilakukan 5 kali pengujian, maka akan terus dilakukan pengujian data dengan menggunakan parameter-parameter yang ada.

#### Referensi

- [1] Dr. Solomon's Virus Encyclopedia, 1995, ISBN:1-897661-00-2, Abstract at <http://vx.netlux.org/lib/aas10.html> (diakses tanggal 2 januari 2010) <http://arykoerniawan.blogspot.co.id/2010/12/mengenal-dan-cara-mengatasi-virus-di.html> (diakses tanggal 5 Mei 2016)
- [2] Anugroho, Prasetyo., dkk. Klasifikasi Email Spam Dengan Metode Naive Bayes Classifier Menggunakan Java Programing.Tugas Akhir. Politeknik Elektronika Negeri Surabaya.2008.
- [3] Kang, Hyohyeong and Choi, Seungjin. Bayesian common spatial patterns for multi-subject EEG classification. Neural Networks, Volume 57 (2014) Pages 39-50
- [4] Mersova, A.,et all. The differentiation of malignant and benign human breast tissue at surgical margins and biopsy using x-ray interaction data and Bayesian classification. Radiation Physics and Chemistry, Volume 95 (2014) 210-213
- [5] Chaphalkara, N.B, et all. Prediction of outcome of construction dispute claims using multilayer perceptron neural network model N.B.International Journal of Project Management, Volume 33,

Issue 8 (2015) 1827-1835

- [6] Orhana, Umut, et all. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications* Volume 38, Issue 10 (2011) 13475–13481.
- [7] Heidari, Elham.,et all. Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN). *Chemometrics and Intelligent Laboratory Systems*, Volume 155(2016)73-85
- [8] Dudek, Grzegorz Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, Volume 32 (2016) 1057-1060
- [9] Fan, Xinghua., el all. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. *Expert Systems with Applications*, Volume 42, Issue 8 (2015) 3945-3952
- [10] Isa, Nor Ashidi Mat., et all. Clustered-Hybrid Multilayer Perceptron network for pattern recognition application. *Applied Soft Computing*, Volume 11 (2011) 1457-1466
- [11] Taner, Danisman., et all. Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron. *Signal Processing*, Volume 93, Issue 6, (2013) 1547-1556