

SELEKSI ATRIBUT PADA ALGORITMA RADIAL BASIS FUNCTION NEURAL NETWORK MENGGUNAKAN INFORMATION GAIN

Yumnah Fitriyanna Waruwu¹, Muhammad Zarlis², Erna Budhiarti Nababan³, Meyman Sokhi Ziliwu⁴

^{1,4}Mahasiswa Pascasarjana Teknik Informatika, Fasilkom Universitas Sumatera Utara

^{2,3}Program Studi Teknik Informatika, Universitas Sumatera Utara

email : ¹yumnahfw@gmail.com

Abstrak: Salah satu metode jaringan syaraf tiruan yang sering digunakan untuk klasifikasi data adalah jaringan RBF karena arsitektur yang sederhana dan pembelajaran jaringan yang cepat. Klasifikasi data dilakukan untuk menentukan kelas pada suatu objek. Sebelum data diklasifikasi, terdapat beberapa atribut yang mempengaruhi keberadaan kelas suatu objek. Beberapa atribut yang terdapat dalam suatu dataset, keberadaannya tidak relevan dan dapat mempengaruhi kinerja jaringan RBF, sehingga berpengaruh pada tingkat akurasi. Oleh sebab itu, perlu dilakukan seleksi atribut agar atribut yang tidak relevan bisa dihilangkan. Pada penelitian ini dilakukan proses seleksi atribut pada sebuah dataset menggunakan information gain. Pada tahap klasifikasi data menggunakan jaringan RBF, hasil akurasi klasifikasi sebelum dilakukan seleksi atribut sebesar 96,59%. Sedangkan setelah proses menghilangkan beberapa atribut, percobaan dilakukan sebanyak 3 kali dengan menghilangkan 1, 2, dan 3 atribut berdasarkan nilai information gain. Dari hasil pengolahan data untuk klasifikasi setelah atribut dihilangkan, rata-rata tingkat akurasi sebesar 96,97%.

Kata Kunci: RBF, klasifikasi, seleksi atribut, information gain, dan tingkat akurasi

PENDAHULUAN

Jaringan saraf tiruan merupakan salah satu metode dalam kecerdasan buatan yang berupaya untuk menirukan kecerdasan manusia dan diterapkan pada mesin atau komputer. Jaringan saraf tiruan bekerja layaknya manusia untuk pekerjaan-pekerjaan tertentu. Menurut Haykin (1994) jaringan saraf tiruan menyerupai otak manusia dalam dua hal, yaitu: pengetahuan diperoleh jaringan melalui proses; dan kekuatan hubungan antara sel syaraf (neuron) yang dikenal sebagai bobot-bobot sinaptik digunakan untuk menyimpan pengetahuan.

Ada banyak kasus yang dapat diselesaikan dengan metode jaringan saraf tiruan seperti: prediksi, *cluster* data, pengenalan pola, klasifikasi, dan lain-lain. Metode ini diterapkan sesuai kasus yang ingin diselesaikan. Banyak penelitian yang telah dilakukan sebelumnya, terutama dalam penyelesaian kasus klasifikasi data. Klasifikasi data merupakan suatu metode yang digunakan untuk menentukan kelas pada suatu objek. Dengan cara ini kita dapat membedakan antara satu objek dengan yang lainnya dan dikelompokkan ke dalam kelas yang berbeda. Salah satu metode yang banyak digunakan untuk klasifikasi data adalah *radial basis function* (RBF) *neural network*.

Jaringan RBF merupakan salah satu jaringan yang cukup populer digunakan karena beberapa alasan, yaitu: struktur jaringan yang sederhana, pembelajaran yang cepat, dan kehandalannya mengklasifikasi data untuk beberapa kasus (Munnoli & Bapat, 2013). Sebagai contoh

penelitian yang dilakukan oleh Pangestuti, et. al (2016) yang melakukan pengklasifikasian kendaraan roda empat dengan ekstraksi ciri *hybrid* berbasis jaringan saraf tiruan *radial basis function*. Pada penelitian ini diperoleh hasil klasifikasi yang memuaskan dengan tingkat akurasi sebesar 77,52%.

Pada kasus pengklasifikasian data pasti dibutuhkan atribut atau ciri yang berguna untuk menentukan objek ke dalam suatu kelas. Banyaknya atribut pada suatu data akan menentukan hasil klasifikasi data. Pada beberapa kasus, atribut yang sangat banyak akan memperlambat kinerja komputer dan tidak jarang ditemukan beberapa atribut yang keberadaannya dirasa kurang bermanfaat bahkan tidak dibutuhkan dan perlu dilakukan seleksi atribut. Menurut Jiawei et. al (2012) seleksi atribut merupakan suatu proses untuk menghilangkan jumlah atribut karena keberadaannya dianggap kurang penting.

Beberapa peneliti telah melakukan pengujian untuk menghilangkan atribut pada suatu data sebelum dan sesudah tahap klasifikasi. Salah satu penelitian yang dilakukan oleh Azhagusundari & Thanamani (2013), yang melakukan penelitian tentang seleksi atribut dengan *information gain* dan memperoleh hasil yang memuaskan. Penelitian ini dilakukan dengan menghilangkan 3 buah atribut pada dataset menggunakan *discernibility matrix* dan *information gain* dengan memperoleh hasil akurasi sebesar 93,6%. Hasil tersebut menunjukkan peningkatan yang baik sebelum dilakukan penyeleksian atribut dengan akurasi sebesar 62,9%. Penelitian lainnya dilakukan oleh Mongkareng, et. al (2017) yang melakukan seleksi fitur pada data

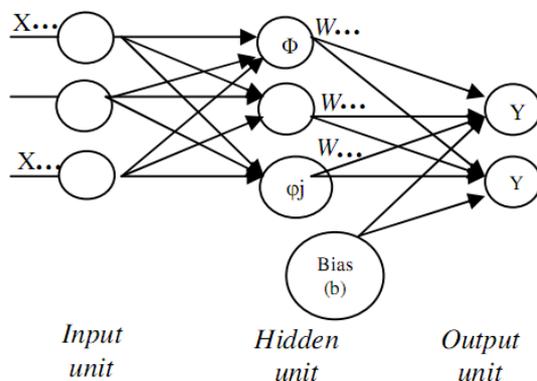
mining. Pada penelitian ini, hasil sebelum dan sesudah penghilangan atribut memperoleh hasil yang berbeda. Pada saat sebelum terjadinya seleksi atribut memperoleh tingkat akurasi sebesar 97% dan setelah seleksi fitur memperoleh hasil sebesar 98%.

Dari kedua penelitian di atas dapat disimpulkan bahwa proses seleksi fitur mempengaruhi hasil akurasi pada klasifikasi data. Beberapa permasalahan umum yang ditemukan sebelum dilakukan pengujian data adalah jumlah atribut yang terkandung. Pada banyak kasus penelitian, beberapa atribut yang mencirikan suatu data tidak relevan bahkan mengurangi tingkat akurasi. Oleh karena itu, pada penelitian ini akan dilakukan proses seleksi atribut pada algoritma jaringan RBF menggunakan *information gain* untuk mengetahui hasil akurasi sebelum dan sesudah menghilangkan atau menyeleksi atribut pada dataset.

METODOLOGI

1. Radial Basis Function

Jaringan saraf tiruan *radial basis function* merupakan salah satu contoh jaringan saraf tiruan dengan metode pelatihan hibrida, yaitu menggabungkan metode pelatihan terbimbing dan metode pelatihan tidak terbimbing. Jaringan RBF dapat digunakan untuk menangani beberapa kasus, terutama untuk pengklasifikasian data. Seperti halnya jaringan saraf tiruan yang lain, Jaringan RBF juga memiliki topologi jaringan. Topologi milik RBF terdiri atas unit lapisan masukan (x), unit lapisan tersembunyi (ϕ), dan unit lapisan keluaran (y). Topologi Jaringan RBF digambarkan sebagai berikut:



Gambar 1: Topologi Jaringan Radial Basis Function

Adapun penjelasan struktur jaringan saraf tiruan *radial basis function*, antara lain:

a. Input layer

Input layer merupakan masukan data dari

jaringan RBF yang terletak pada lapisan pertama. *Input* pada jaringan RBF selain digunakan sebagai data *input* juga berfungsi untuk menentukan *centroid* dan standar deviasi dari data yang berfungsi membangkitkan fungsi aktivasi pada *hidden layer*. Terdapat dua cara yang digunakan dalam penentuan *centroid* yaitu dengan metode acak dan menggunakan algoritma *k-means clustering*. Pada penelitian ini untuk memperoleh nilai *centroid* dilakukan dengan cara menggunakan algoritma *k-means clustering*.

Setelah proses pencarian nilai *centroid* dari data *input* menggunakan algoritma *k-means clustering* dilakukan, tahap selanjutnya dilakukan perhitungan nilai *Gaussian* dengan menggunakan persamaan *Gaussian*. Nilai *centroid* hasil pencarian dari data *input* digunakan pada *hidden layer*.

b. Hidden layer

Pada lapisan ini digunakan fungsi aktivasi untuk memperoleh bobot jaringan yang digunakan pada *output layer*. Penerapan fungsi yang umum digunakan dalam pemecahan kasus menggunakan jaringan RBF yaitu dengan menggunakan fungsi basis *Gaussian* yang diformulasikan sebagai berikut:

$$\varphi(r) = \exp\left(-\frac{\|x_i - c_i\|^2}{2\sigma_i^2}\right) \quad (1)$$

Dengan:

$$\sigma = \frac{\text{Jarak maksimum antara 2 pusat}}{\sqrt{\text{banyak pusat}}} = \frac{d_{\max}}{\sqrt{n}}$$

Setelah nilai *Gaussian* diketahui, maka proses selanjutnya adalah menghitung nilai bobot baru (w) dengan mengalikan *pseudoinverse* dari matriks G (*Gaussian*) dengan vector target (d) dengan persamaan berikut:

$$w = (G^T G)^{-1} G^T d \quad (2)$$

Setelah nilai *weight* diketahui maka tahap selanjutnya adalah menyimpan nilai hasil perhitungan *centroid* dan nilai *weight* untuk dapat digunakan kembali pada saat proses pengujian pada jaringan RBF *neural network*.

c. Output Layer

Output layer merupakan langkah terakhir setelah memperoleh hasil perhitungan dari fungsi basis *Gaussian* untuk melakukan pembaharuan bobot. Penjumlahan dari perkalian antara bobot dengan fungsi basis akan menghasilkan keluaran yang disebut *output layer*. *Output layer* merespon dari jaringan sesuai pola yang dimasukkan sebelumnya pada *input layer*. Transformasi dari ruang masukan ke ruang *hidden layer* adalah *non-linier*, sedangkan transformasi dari ruang *hidden*

layer ke *output layer* adalah linier. Pada tahap ini dilakukan untuk menghitung nilai *output* jaringan $y(n)$ ditambah dengan bobot bias (b) dengan persamaan berikut:

$$y(n) = \sum_{i=1} wG(\|x - c\| + b)$$

Menurut Munnoli & Bapat (2013) beberapa kelebihan penggunaan jaringan RBF, antara lain :

1. Jaringan RBF bersifat *feed-forward*.
2. Pada *hidden layer* digunakan sebuah fungsi aktivasi yang berbasis *radial*, misalnya fungsi *Gaussian*.
3. Pada *output unit* menggunakan penjumlahan secara *linier*.
4. Pada data pelatihan, langkah dibagi menjadi dua: pertama, bobot ditentukan dari *input* ke *hidden layer* secara *nonlinier* dan kemudian bobot dari *hidden layer* ke *output layer* dilakukan secara *linier*.
5. Pelatihan jaringan sangat cepat.
6. Sangat baik digunakan untuk kasus interpolasi.

2. Information Gain

Information Gain merupakan metode seleksi fitur paling sederhana dengan melakukan perbandingan atribut dan banyak digunakan dalam aplikasi kategorisasi teks, analisis data microarray dan analisis data citra. Penentuan atribut terbaik dilakukan dengan menghitung nilai *entropy* terlebih dahulu. *Entropy* merupakan ukuran ketidakpastian kelas dengan menggunakan *probabilitas* kejadian atau atribut tertentu. Berikut ini rumus entropi

$$E(Y) = -\sum_{y \in Y} p(y) \log_2(p(y)) \quad (3)$$

Kemudian entropi y setelah mengamati x adalah :

$$E(Y|X) = -\sum_{y \in Y} p(x) \sum_{y \in Y} p(y|x) \quad (4)$$

dimana $p(y|x)$ sama dengan probabilitas bersyarat dari y terhadap x . Mengingat entropi sebagai kriteria pengotor dalam pelatihan set S , kita dapat mendefinisikan ukuran mencerminkan informasi tambahan tentang Y disediakan oleh X yang mewakili jumlah dimana entropi Y menurun. Langkah ini dikenal sebagai IG. Hal ini diberikan oleh :

$$IG = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad (5)$$

Seperti yang terlihat pada persamaan di atas bahwa informasi yang diperoleh tentang Y setelah mengamati X sama dengan informasi yang diperoleh tentang X setelah mengamati Y .

HASIL DAN PEMBAHASAN

Setelah dilakukan proses training, testing, dan seleksi atribut dari data set wine maka proses pengujian dilakukan sebanyak 4 kali yang dimulai dari tabel percobaan 1 di bawah ini :

Percobaan 1		
No	Selected Attribute	Akurasi Pengujian (%)
1	All	96,59
2	3 (ash)	96,59

Dari tabel percobaan 1, akurasi pengujian apabila seleksi atribut tidak ada direduksi menghasilkan akurasi pengujian sebesar 96,59 % dan apabila salah satu atribut direduksi yaitu atribut ke-3 yang direduksi menghasilkan akurasi pengujian yang sama yaitu sebesar 96,59 %. Atribut ke-3 yang direduksi dikarenakan atribut ini menghasilkan nilai information gain terkecil. Untuk tabel percobaan ke-2 dapat dilihat di bawah ini

Percobaan 2		
No	Selected Attribute	Akurasi Pengujian (%)
1	All	96,59
2	1 (alcohol)	94,31

Untuk tabel percobaan ke-2, sama halnya dengan percobaan ke-1 yaitu dengan membandingkan antara atribut tanpa reduksi dan salah satu atribut yang direduksi yaitu atribut 1. Hasil akurasi pengujian ini sendiri menghasilkan akurasi pengujian yang berbeda dimana hasil akurasi tanpa direduksi menghasilkan akurasi pengujian sebesar 96,59 % dan atribut 1 direduksi yaitu 94,31 %. Atribut 1 dipilih untuk direduksi dikarenakan atribut ini memiliki nilai information gain terbesar. Kemudian akan dilanjutkan dengan tabel percobaan 3 seperti di bawah ini.

Percobaan 3		
No	Selected Attribute	Akurasi Pengujian (%)
1	All	96,59
2	3 (Ash) dan 5 (magnesium)	96,59

Dari percobaan 3, yang dibandingkan adalah semua atribut dan 2 atribut yang dihilangkan. Hasil akurasi pengujian dari proses ini menghasilkan nilai akurasi yang sama yaitu 96,55 %. Kedua atribut yang direduksi menghasilkan nilai information 2 terendah. Terakhir akan ditampilkan tabel perbandingan untuk percobaan ke-4 yang ditampilkan pada tabel di bawah ini

Percobaan 4		
No	Selected Attribute	Akurasi Pengujian (%)
1	All	96,59
2	3 (Ash), 4 (alcalinity of ash), dan 5 (magnesium)	97,72

Pada tabel percobaan ke-4 ini ada 3 atribut yang dihilangkan dan akan dibandingkan dengan tanpa menghilangkan atribut. Terlihat pada tabel tersebut jika 3 atribut dihilangkan maka menghasilkan akurasi pengujian yang lebih tinggi yaitu 97,7 % dan tanpa menghilangkan atribut sebesar 96,59.

SIMPULAN

Berdasarkan 4 pengujian perbandingan yang sudah dilakukan, apabila nilai atribut information gain yang terkecil direduksi tidak mempengaruhi hasil prediksi klasifikasi terlihat jelas di tabel percobaan ke-1 dan ke-3 menghasilkan nilai akurasi yang sama dengan tanpa menghilangkan atribut dan bahkan percobaan ke-4 jika 3 atribut terendah yang

dihilangkan menghasilkan nilai akurasi yang lebih tinggi daripada tanpa reduksi. Hal sebaliknya terjadi jika nilai information gain dari atribut tersebut tinggi seperti percobaan ke-2 yang direduksi maka hasil pengujian klasifikasi menghasilkan nilai akurasi yang lebih tinggi jika tanpa reduksi dan mereduksi 1 atribut yang menghasilkan nilai information gain tertinggi. Hal ini membuktikan bahwa apabila proses seleksi atribut sangat berfungsi untuk melakukan proses klasifikasi yang dimana dapat menseleksi atribut yang tidak terlalu berdampak dalam proses klasifikasi dan dapat menghasilkan hasil klasifikasi yang lebih cepat secara waktu.

DAFTAR PUSTAKA

- Azhagusundari, B. & Thanamani, A.S. 2013. Feature Selection based on Information Gain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)- Volume-2, Issue-2*. Online. <https://pdfs.semanticscholar.org>
- Haykin, S. (1994). Neural networks-A comprehensive foundation. *New York: IEEE Press. Herrmann, M., Bauer, H.-U., & Der, R.* <https://doi.org/10.1017/S0269888998214044>
- Mongkareng, D., Setiawan, N.A., Permanasari, A.E. 2017. Implementasi Data Mining dengan Seleksi Fitur untuk Klasifikasi Serangan pada Intrusion Detection System (IDS). *CITEE*. Online. citee.ft.ugm-ac.id/download51.php
- Munnoli, A.P. & Bapat, A.U. (2013). Clustering Algorithms for Radial Basis Function Neural Network. *ITSI Transactions on Electrical and Electronics Engineering (ITSI-TEEE) - Volume -1, Issue*. (Online). <https://pdfs.semanticscholar.org>
- Jiawei, H., Kamber, M. & Pei, J. 2012. *Data Mining: Concepts and Techniques (Third Edition)*. Elsevier : Wyman Street, Waltham, MA 02451, USA
- Pangestuti, G.W., Usman, K. & Purnama, B. 2016. Klasifikasi Kendaraan Roda Empat dengan Ekstraksi Ciri Hybrid Berbasis Jaringan Syaraf Tiruan. *e-Proceeding of Engineering : Vol.3, No.2 Agustus 2016*. (online).