

PERBANDINGAN ALGORITMA C4.5 DAN NEURAL NETWORK UNTUK MEMPREDIKSI HASIL PEMILU LEGISLATIF DKI JAKARTA

Mohammad Badrul

Program Studi Sistem Informasi
STMIK Nusa Mandiri Jakarta

Jl. Damai No. 8 Warung Jati Barat (Margasatwa) Jakarta Selatan. Telp. (021) 78839513
mohammad.mbl@nusamandiri.ac.id

ABSTRACT

General elections are a means of implementation of the sovereignty of the people in the Unitary State of Indonesia based on Pancasila and 1945 Constitution. Elections held in Indonesia is to choose the leadership of both the president and vice president, member of parliament, parliament, and the DPD. In this study comparison of data mining methods, namely C4.5 and neural network algorithm is applied to both the data legislative candidates to be elected to the legislature and were not selected. C4.5 algorithm is one of the algorithms in a decision tree method that converts the data into a decision tree using entropy calculation formula. While the neural network algorithm is a method like human neurons to find the best path. From the test results to measure the performance of both methods using cross validation test method, confusion matrix and ROC curves is known that the neural network has the highest accuracy value which is equal to 98.50%, followed by the C4.5 algorithm method with 97.84% accuracy values. AUC values for the neural network method showed the highest value of 0.982 and a decision tree algorithm with a value of 0.970.

Kata Kunci: Elections, C4.5 algorithm, neural network algorithm, accuracy.

PENDAHULUAN

Pemilihan umum (Pemilu) adalah sarana pelaksanaan kedaulatan rakyat dalam negara kesatuan Republik Indonesia yang berdasarkan Pancasila dan Undang-Undang Dasar 1945 (Undang-Undang RI No.10 Tahun 2008 tentang Pemilihan umum). Pemilihan umum adalah salah satu pilar utama untuk memilih pemimpin dari sebuah demokrasi atau bisa disebut yang terutama (Santoso, 2004). Pemilu merupakan sarana yang sangat penting bagi terselenggaranya sistem politik yang demokratis. Karena itu, tidak mengherankan banyak negara yang ingin disebut sebagai negara demokratis menggunakan pemilu sebagai mekanisme membangun legitimasi. Pemilu bertujuan untuk memilih anggota Dewan Perwakilan Rakyat (DPR), Dewan Perwakilan Rakyat Daerah (DPRD) provinsi, dan Dewan Perwakilan Rakyat Daerah (DPRD) kabupaten/kota yang dilaksanakan dengan sistem proporsional terbuka (Undang-Undang RI No.10 Tahun 2008 tentang Pemilihan umum). Dengan sistem pemilu langsung dan jumlah partai yang besar maka pemilu legislatif memberikan peluang yang besar pula bagi rakyat Indonesia untuk berkompetisi menaikkan diri menjadi anggota

legislatif. Pemilu legislatif tahun 2009 diikuti sebanyak 44 partai yang terdiri dari partai nasional dan partai lokal. Pemilu Legislatif DKI Jakarta Tahun 2009 terdapat 2.268 calon anggota Dewan Perwakilan Rakyat Daerah (DPRD) dari 44 partai yang akan bersaing memperebutkan 94 kursi anggota Dewan Perwakilan Rakyat DKI Jakarta.

Prediksi hasil pemilihan umum perlu diprediksi dengan akurat, karena hasil prediksi yang akurat sangat penting karena mempunyai dampak pada berbagai macam aspek sosial, ekonomi, keamanan, dan lain-lain (Borisjuk, Borisjuk, Rallings, dan Thrasher, 2005). Bagi para pelaku ekonomi, peristiwa politik seperti pemilu tidak dapat dipandang sebelah mata, mengingat hal tersebut dapat mengakibatkan risiko positif maupun negatif terhadap kelangsungan usaha yang dijalankan.

Metode prediksi hasil pemilihan umum sudah pernah dilakukan oleh peneliti (Rigdon, Jacobson, Sewell, dan Rigdon, 2009) melakukan prediksi hasil pemilihan umum dengan menggunakan metode *Estimator Bayesian*. (Moscato, Mathieson, Mendes, dan Berreta, 2005) melakukan penelitian untuk memprediksi pemilihan presiden Amerika Serikat menggunakan *decision tree*. (Choi dan Han, 1999)

memprediksi hasil pemilihan presiden di Korea dengan metode Decision Tree. (Nagadevara dan Vishnuprasad, 2005) memprediksi hasil pemilihan umum dengan model *classification tree dan neural network*. (Borisyyuk, Borisyyuk, Rallings, dan Thrasher, 2005) yang memprediksi hasil pemilihan umum dengan menggunakan metode *neural network*. Tetapi belum ada yang melakukan perbandingan metode-metode tersebut sehingga belum diketahui metode yang paling akurat.

Untuk itu dalam penelitian ini akan dilakukan perbandingan metode *neural network dengan* metode yang layaknya neuron manusia untuk mencari jalur terbaik dan metode algoritma C4.5 dengan merubah data menjadi pohon keputusan menggunakan rumus perhitungan entropi. Pada penelitian ini, penulis menggunakan aplikasi rapidminer sehingga diperoleh metode dengan akurasi prediksi hasil pemilihan umum Legislatif di DKI Jakarta.

BAHAN DAN METODE

Pemilihan umum (Pemilu) adalah salah satu pilar utama dari sebuah demokrasi, kalau tidak dapat yang disebut yang terutama. Sentralitas dari posisi pemilihan umum dalam membedakan sistem politik yang demokratis dan bukan tampak jelas dari beberapa definisi yang diajukan oleh beberapa peneliti. Salah satu konsepsi modern awal mengenai demokrasi yang diajukan oleh Joseph Schumpeter dan kemudian dikenal dengan mazhab Schumpeterian menempatkan penyelenggaraan pemilihan umum yang bebas dan berkala sebagai kriteria utama bagi sebuah sistem politik untuk dapat disebut sebagai sebuah demokrasi (Santoso, 2004). Dalam sebuah negara demokrasi, pemilihan umum merupakan salah satu pilar utama untuk memilih pemimpin yang nantinya akan mewakili rakyat untuk duduk dipemerintahan mulai dari anggota Dewan Perwakilan Rakyat Daerah (DPRD) tingkat II, Dewan Perwakilan Rakyat Daerah (DPRD) Tingkat I, Dewan Perwakilan Rakyat (DPR) RI dan Dewan Pimpinan Daerah (DPD).

Dalam khazanah demokrasi kontemporer, posisi pemilihan umum memperoleh penguatan. Kajian akademis mengenai demokrasi mengenal dua kategori pemaknaan besar, yaitu konsepsi minimalis dan maksimalis. Demokrasi minimalis atau prosedural dikenakan kepada sistem-sistem politik yang melaksanakan perubahan kepemimpinan secara reguler melalui suatu

mekanisme pemilihan yang berlangsung bebas, terbuka dan melibatkan massa pemilih yang universal. Sedangkan konsep maksimalis adalah pelaksanaan pemilihan umum tidaklah cukup bagi suatu sistem politik untuk mendapatkan gelar demokrasi karena konsep ini mensyaratkan penghormatan terhadap hak-hak sipil yang lebih luas (Santoso, 2004). Pemilu di Indonesia terbagi dari dua bagian, yaitu (Sardini, 2011):

1. Pemilu orde baru

Sistem pemilihannya dilakukan secara proporsional tidak murni, yang artinya jumlah penentuan kursi tidak ditentukan oleh jumlah penduduk saja tetapi juga didasarkan pada wilayah administrasi. Pemilu orde baru dimulai pada tahun 1955 sebagai pemilu pertama yang diselenggarakan di negara Indonesia.

2. Pemilu era reformasi

Dikatakan sebagai pemilu reformasi karena dipercepatnya proses pemilu di tahun 1999 sebelum habis masa kepemimpinan di pemilu tahun 1997. Terjadinya pemilu era reformasi ini karena produk pemilu pada tahun 1997 dianggap pemerintah dan lembaga lainnya tidak dapat dipercaya.

Sistem pemilihan Dewan Perwakilan Rakyat (DPR)/Dewan Perwakilan Rakyat Daerah (DPRD) berdasarkan ketentuan dalam Undang-Undang nomor 10 tahun 2008 pasal 5 ayat 1 sistem yang digunakan dalam pemilihan legislatif adalah sistem proporsional dengan daftar terbuka, sistem pemilihan DPD dilaksanakan dengan sistem distrik berwakil banyak Undang-Undang nomor 10 tahun 2008 pasal 5 ayat 2 tentang pemilihan umum. Menurut Undang-Undang No. 10 tahun 2008 tentang pemilihan umum, Peserta pemilihan anggota DPR/D adalah partai politik peserta Pemilu, sedangkan peserta pemilihan anggota DPD adalah perseorangan. Partai politik peserta Pemilu dapat mengajukan calon sebanyak-banyaknya 120 persen dari jumlah kursi yang diperebutkan pada setiap daerah pemilihan demokratis dan terbuka serta dapat mengajukan calon dengan memperhatikan keterwakilan perempuan sekurang-kurangnya 30 %. Partai Politik Peserta Pemilu diharuskan Undang-Undang untuk mengajukan daftar calon dengan nomor urut (untuk mendapatkan Kursi). Karena itu dari segi pencalonan Undang-Undang No.10 Tahun 2008 tentang Pemilihan umum mengadopsi sistem daftar calon tertutup.

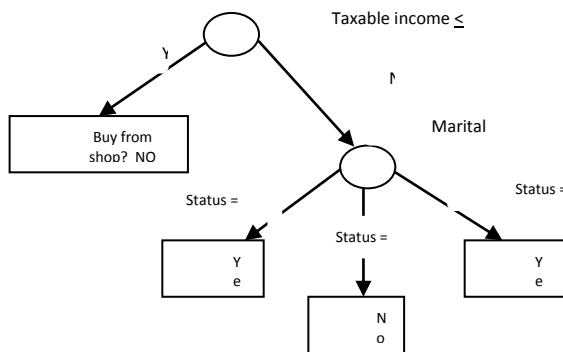
Undang-Undang No.10 Tahun 2008 tentang Pemilihan Umum mengadopsi sistem

proporsional dengan daftar terbuka. sistem proporsional merujuk pada formula pembagian kursi dan/atau penentuan calon terpilih, yaitu setiap partai politik peserta pemilu mendapatkan kursi proporsional dengan jumlah suara sah yang diperolehnya. Penerapan formula proporsional dimulai dengan menghitung Bilangan Pembagi Pemilih (BPP), yaitu jumlah keseluruhan suara sah yang diperoleh seluruh partai politik peserta pemilu pada suatu daerah pemilihan dibagi dengan jumlah kursi yang diperebutkan pada daerah pemilihan tersebut.

Algoritma C4.5

Algoritma C4.5 adalah hasil dari pengembangan algoritma ID3 (*Iterative Dichotomiser*) yang dikembangkan oleh Quinlan (Han dan Kamber 2006). Algoritma ini digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (Gorunescu, 2011). Sebelumnya diakhir tahun 1970 sampai awal tahun 1980 J. Ross Quinlan, seorang peneliti dibidang *machine learning*, membuat sebuah algoritma *decision tree* yang dikenal dengan ID3 (*Iterative Dichotomiser*).

Algoritma C4.5 atau pohon keputusan mirip sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan dengan mudah dapat dikonversi ke aturan klasifikasi. Secara umum keputusan pengklasifikasi pohon memiliki akurasi yang baik, namun keberhasilan penggunaan tergantung pada data yang akan diolah.



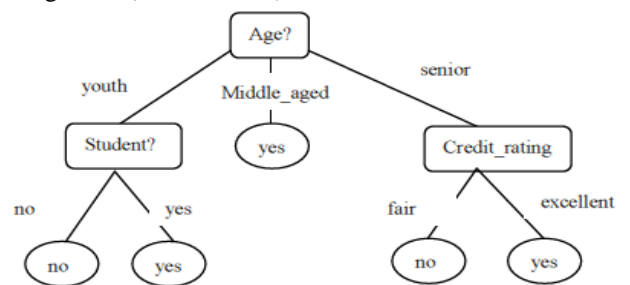
Sumber : Gorunescu, 2011

Gambar 1 Contoh konsep pohon keputusan sederhana

Pada Gambar 1 variabel target untuk pohon keputusan adalah membeli pada toko,

dengan pengklasifikasian ya atau tidak. Variabel *predictor* adalah taxable income (≤ 34.000 atau > 34.000), marital status (married, divorced, atau single). Simpul akar merupakan simpul keputusan, pengujiannya apakah taxable income ≤ 34.000 atau > 34.000 .

Desicion Tree menyerupai struktur *flowchart*, yang masing-masing internal *node*-nya dinyatakan sebagai atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap *node* daun (*terminal node*) menentukan label *class*. *Node* paling atas dari sebuah pohon adalah *node* akar (Han dan Kamber, 2007). Salah satu metode klasifikasi yang menarik melibatkan konstruksi pohon keputusan, koleksi *node* keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di *node* daun. Dimulai di *node root*, yang oleh konvensi ditempatkan di bagian atas dari diagram pohon keputusan, atribut diuji pada *node* keputusan, dengan setiap hasil yang mungkin dihasilkan dalam suatu cabang. Setiap cabang kemudian mengarah baik ke *node* lain keputusan atau ke *node* daun untuk mengakhiri (Larose, 2005).



Sumber : Larose, 2005

Gambar 2. Contoh Decision Tree

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (kusrini dan Lutfi, 2009), yaitu:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy yaitu:

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

Keterangan:

S : himpunan kasus

A : atribut

N : jumlah partisi

Pi : proporsi dari Si terhadap S

3. Kemudian hitung nilai gain dengan metode *informasi gain*:

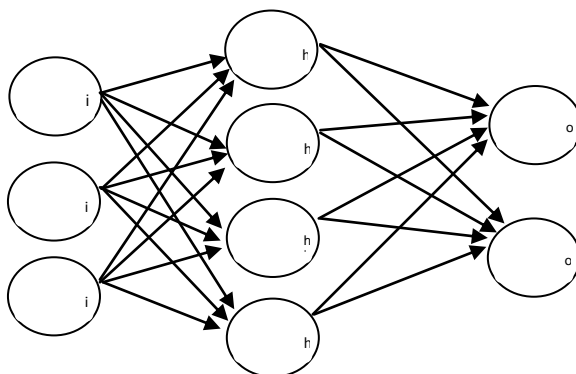
$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

4. Ulangi langkah ke-2 hingga semua tupel terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua tupel dalam node N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam tupel yang dipartisi lagi.
 - c. Tidak ada tupel di dalam cabang yang kosong.

Neural Network

Neural network adalah suatu sistem pemroses informasi yang memiliki karakteristik menyerupai dengan jaringan saraf biologi pada manusia. *Neural network* didefinisikan sebagai sistem komputasi di mana arsitektur dan operasi diilhami dari pengetahuan tentang sel saraf biologis di dalam otak, yang merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba menstimulasi proses pembelajaran pada otak manusia tersebut (Astuti, 2009).

Neural network dibuat berdasarkan model saraf manusia tetapi dengan bagian-bagian yang lebih sederhana. Komponen terkecil dari *neural network* adalah unit atau yang biasa disebut dengan *neuron* dimana *neuron* tersebut akan mentransformasikan informasi yang diterima menuju *neuron* lainnya (Shukla, Tiwari, dan Kala, 2010).



Sumber : Shukla, Tiwari, dan Kala, 2010
Gambar 3 Model *neural network*

Neural network terdiri dari dua atau lebih lapisan, meskipun sebagian besar jaringan terdiri dari tiga lapisan: lapisan input, lapisan tersembunyi, dan lapisan output (Larose, 2005). Pendekatan *neural network* dimotivasi oleh jaringan saraf biologis. Secara kasar, *neural network* adalah satu set terhubung *input/output* unit, di mana masing-masing sambungan memiliki berat yang terkait dengannya. *Neural network* memiliki beberapa properti yang membuat mereka populer untuk *clustering*. Pertama, *neural network* adalah arsitektur pengolahan *inherent* paralel dan terdistribusi. Kedua, *neural network* belajar dengan menyesuaikan bobot interkoneksi dengan data, Hal ini memungkinkan *neural network* untuk "menormalkan" pola dan bertindak sebagai fitur (atribut) *extractors* untuk kelompok yang berbeda. Ketiga, *neural network* memproses vektor numerik dan membutuhkan pola objek untuk diwakili oleh fitur kuantitatif saja (Gorunescu, 2011).

Neural network terdiri dari kumpulan node (*neuron*) dan relasi. Ada tiga tipe node (*neuron*) yaitu, *input*, *hidden* dan *output*. Setiap relasi menghubungkan dua buah node dengan bobot tertentu dan juga terdapat arah yang menunjukkan aliran data dalam proses (Kusrini dan Luthfi, 2009). Kemampuan otak manusia seperti mengingat, menghitung, mengeneralisasi, adaptasi, diharapkan *neural network* dapat meniru kemampuan otak manusia. *Neural network* berusaha meniru struktur/arsitektur dan cara kerja otak manusia sehingga diharapkan bisa dan mampu menggantikan beberapa pekerjaan manusia. *Neural network* berguna untuk memecahkan persoalan yang berkaitan dengan pengenalan pola, klasifikasi, prediksi dan data mining (Shukla, Tiwari, dan Kala, 2010).

Input node terdapat pada *layer* pertama dalam *neural network*. Secara umum setiap input node merepresentasikan sebuah input parameter seperti umur, jenis kelamin, atau pendapatan. *Hidden node* merupakan node yang terdapat di bagian tengah. *Hidden node* ini menerima masukan dari input node pada *layer* pertama atau dari *hidden node* dari *layer* sebelumnya. *Hidden node* mengombinasikan semua masukan berdasarkan bobot dari relasi yang terhubung, mengkalkulasikan, dan memberikan keluaran untuk *layer* berikutnya. *Output node* mempresentasikan atribut yang diprediksi (Kusrini dan Luthfi, 2009).

Setiap node (*neuron*) dalam *neural network* merupakan sebuah unit pemrosesan.

Tiap node memiliki beberapa masukan dan sebuah keluaran. Setiap node mengkombinasikan beberapa nilai masukan, melakukan kalkulasi, dan membangkitkan nilai keluaran (aktifasi). Dalam setiap node terdapat dua fungsi, yaitu fungsi untuk mengkombinasikan masukan dan fungsi aktifasi untuk menghitung keluaran. Terdapat beberapa metode untuk mengkombinasikan masukan antara lain *weighted sum*, *mean*, *max*, logika OR, atau logika AND (Kusrini dan Luthfi, 2009). Serta beberapa fungsi aktifasi yang dapat digunakan yaitu *heaviside (threshold)*, *step activation*, *piecewise*, *linear*, *gaussian*, *sigmoid*, *hyperbolic tangent* (Gorunescu, 2011).

Salah satu keuntungan menggunakan *neural network* adalah bahwa *neural network* cukup kuat sehubungan dengan data. Karena *neural network* berisi banyak node (*neuron* buatan) dengan bobot ditugaskan untuk setiap koneksi (Larose, 2005).

Algoritma *neural network* mempunyai karakteristik-karakteristik lainnya antara lain (Astuti, 2009),

1. Masukan dapat berupa nilai diskrit atau real yang memiliki banyak dimensi..
2. Keluaran berupa *vektor* yang terdiri dari beberapa nilai diskrit atau real
3. Dapat mengetahui permasalahan secara *black box*, dengan hanya mengetahui nilai masukan serta keluarannya saja.
4. Mampu menangani pembelajaran terhadap data yang memiliki derau (*noise*).
5. Bentuk dari fungsi target pembelajaran tidak diketahui karena hanya berupa bobot-bobot nilai masukan pada setiap *neuron*.
6. Karena harus mengubah banyak nilai bobot pada proses pembelajaran, maka waktu pembelajaran menjadi lama, sehingga tidak cocok untuk masalah-masalah yang memerlukan waktu cepat dalam pembelajaran.
7. *Neural network* hasil pembelajaran tiruan dapat dijalankan dengan tepat.

Penelitian di bidang *neural network* dimulai pada masa komputer digital. McCulloch dan Pitts (1943) mengemukakan model matematika pertama untuk *neural network*. Rosenblatt (1962) mengemukakan model *perceptron* dan algoritma pembelajaran pada tahun 1962 (Alpaydin, 2010). Minsky dan Papert (1969) menunjukkan keterbatasan *single layer perceptron* untuk menyelesaikan masalah yang *nonlinearly separable*. Kemudian Rumelhart, Hinton, and Williams (1986)

yang mempresentasikan algoritma *backpropagation* untuk *multilayer perceptron* yang dapat menyelesaikan masalah yang *nonlinearly separable* (Han dan Kamber, 2007). Aplikasi *neural network* telah banyak dimanfaatkan untuk berbagai kepentingan seperti di bidang Elektronik, Otomotif, Perbankan, Sistem penerbangan udara, Dunia hiburan, transportasi publik, telekomunikasi, bidang Kesehatan, Keamanan, bidang Robotika, Asuransi, Pabrik, Financial, Suara, Pertambangan dan sistem pertahanan (Astuti, 2009).

Algoritma yang paling populer pada algoritma *neural network* adalah algoritma *backpropagation*. Algoritma pelatihan *backpropagation* atau ada yang menterjemahkan menjadi propagasi balik pertama kali dirumuskan oleh Paul Werbos pada tahun 1974 dan dipopulerkan oleh Rumelhart bersama McClelland untuk dipakai pada *neural network*. Metode *backpropagation* pada awalnya dirancang untuk *neural network feedforward*, tetapi pada perkembangannya, metode ini diadaptasi untuk pembelajaran pada model *neural network* lainnya (Astuti, 2009). Salah satu metode pelatihan terawasi pada *neural network* adalah metode *backpropagation*, di mana ciri dari metode ini adalah meminimalkan *error* pada output yang dihasilkan oleh jaringan.

Metode algoritma *backpropagation* ini banyak diaplikasikan secara luas. *backpropagation* telah berhasil diaplikasikan di berbagai bidang, antaranya bidang finansial, pengenalan pola tulisan tangan, pengenalan pola suara, sistem kendali, pengolah citra medika. *backpropagation* berhasil menjadi salah satu metode komputasi yang handal. Algoritma *backpropagation* mempunyai pengaturuan hubungan yang sangat sederhana yaitu: jika keluaran memberikan hasil yang salah, maka penimbang (*weight*) dikoreksi supaya galatnya dapat diperkecil dan respon jaringan selanjutnya diharapkan akan mendekati nilai yang benar. Algoritma ini juga berkemampuan untuk memperbaiki penimbang pada lapisan tersembunyi (*hidden layer*) (Purnomo dan Kurniawan, 2006).

Inisialisasi awal bobot jaringan *backpropagation* yang terdiri atas lapisan input, lapisan tersembunyi, dan lapisan output (Astuti, 2009). Tahap pelatihan *backpropagation* merupakan langkah untuk melatih suatu *neural network* yaitu dengan cara melakukan perubahan penimbang (sambungan antar lapis yang membentuk *neural network* melalui masing-masing

