

# Neural Network dan Implementasinya Dalam Data Mining

Rudolf Rudi Hermanto

Institut Teknologi Bandung

rudolf@students.itb.ac.id

## Abstrak

Untuk mendapatkan informasi yang berharga dibutuhkan algoritma-algoritma yang diterapkan dalam *datamining* di database yang besar. Terdapat banyak sekali algoritma algoritma kompleks dalam *datamining*. *Neural network* memiliki peranan dalam *datamining*. *Neural network* bisa dipakai sebagai *benchmark* untuk membangun model *logistic regression* dalam data mining.

## 1. Pendahuluan

Perbedaan data dan informasi bisa dianalogikan dengan perbandingan antara batu dan 99,99% *ingot* emas murni. Kita tahu bahwa ada emas di dalam batu. Tapi kita tidak tahu dimana letak pastinya. Sama halnya dengan data. *Data mining* adalah istilah yang digunakan untuk menggambarkan proses *extract* nilai/informasi dari database. Empat hal dibutuhkan agar data mining berjalan secara efektif: data yang memiliki kualitas yang tinggi, *right data*, contoh yang *adequate*, dan perangkat yang benar. Terdapat banyak algoritma dalam *data mining*. Salah satunya adalah *neural network* sebagai *benchmark* untuk membangun model dalam *datamining*.

## 2. Neural Network

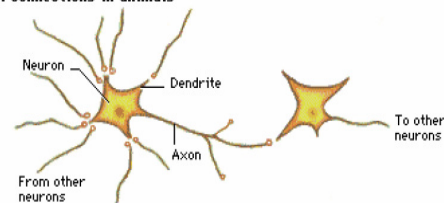
Disebut *neural network* karena didesain mengikuti cara otak memproses dan menyimpan informasi. Otak manusia terdiri dari ratusan sampai jutaan neuron. Setiap neuron memiliki desain yang sederhana.

*Neural network* disusun dari nodes yang mengkombinasikan input-inputnya (variabel dari database atau output dari node yang lain). Node ini dapat diklasifikasikan dari tiga layer yang sederhana. *Input layer*, *output layer* dan *middle layer*.

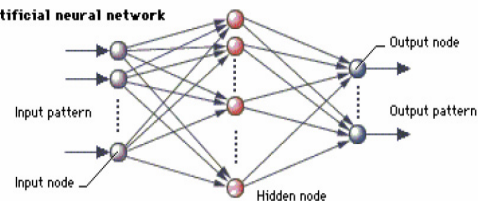
Pertama-tama kita membuat model *logistic* yang sederhana yang menunjukkan *neural network*. Di sini, tidak dibutuhkan waktu yang cukup panjang untuk membangun model yang lebih baik menggunakan *logistic refression* dibandingkan *neural network*.

Dalam prakteknya, *neural network* adalah perangkat pemodelan data statistik non linear. *Neural network* dapat digunakan untuk memodelkan relasi yang kompleks antara input dan output untuk mencari pola dalam data. *Neural network* secara esensial terdiri dari tiga bagian: arsitektur atau model, algoritma learning, dan fungsi aktivasi. Pada gambar di bawah ini diperlihatkan *artificial neural network*.

Neural connections in animals



Artificial neural network



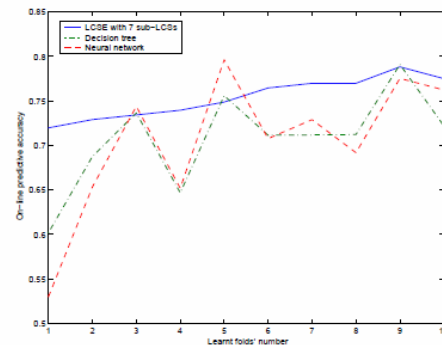
bawah ini diperlihatkan perbandingan LCSE, decision tree, dan neural network.

### 3. Keuntungan Neural Network

1. Akurasi yang tinggi: *Neural network* digunakan untuk *mapping* aproksimasi kompleks non linear.
2. Toleransi terhadap *noise*: *Neural network* sangat fleksibel dengan data yang *noisy*.
3. Independensi dari asumsi *prior*: *Neural network* tidak membuat asumsi priori tentang distribusi data atau bentuk interaksi dari faktor-faktor.
4. Mudah untuk dikelola: *Neural network* dapat diupdate dengan data yang baru, membuat berguna untuk lingkungan yang dinamis.
5. *Neural network* dapat diimplementasikan di hardware yang paralel.
6. Ketika element *neural network* gagal, ia dapat melanjutkan tanpa masalah karena polanya yang paralel.
7. *Neural network* dapat dilatih di dataset yang sangat besar secara iteratif.

### 4. Perbandingan algoritma-algoritma dalam data mining

Terdapat banyak algoritma-algoritma dalam data mining. Algoritma-algoritma yang akan dibandingkan adalah learning classifier system, neural networks dan decision tree. Learning classifier system ensemble (LCSE) mengkombinasikan sistem *learning classifier* dengan *ensemble learning* dalam rangka memperlengkap sistem. Pada gambar di



Comparison between LCSE and Neural Network and Decision Tree.

### 5. Neural Network Dan Implementasinya Dalam Data Mining.

Terdapat faktor-faktor penting dalam implementasi neural network:

- a. Kombinasi yang efektif antara Neural Network dan Teknologi Data Mining

Teknologi umumnya menggunakan software neural network atau mengtransformasikan dari perangkat neural network yang sudah dikembangkan, diagram kerja dari data mining harus dimengerti sebaik-baiknya. Model data dan perangkat muka aplikasi harus bisa dikembangkan dengan form yang sudah distandardsasikan sehingga dua teknologi dapat efektif diintegrasikan dan bersama-sama memiliki peran dalam data mining. Oleh sebab itu dibutuhkan pendekatan yang lebih baik lagi dalam mengkombinasikan neural network dan teknologi data mining harus dapat dikembangkan lebih baik lagi untuk terciptanya teknologi data mining yang lebih baik ke depannya.

- b. Kombinasi yang efektif antara proses knowledge dan komputasi neural.

Mengevaluasi apakah suatu pelaksanaan algoritma data mining baik; indikator berikut dan karakteristiknya dapat digunakan: (1) apakah pemodelan berkualitas tinggi dalam keadaan "noisy" dan data masih setengah matang, (2) model harus dipahami oleh pengguna dan dapat digunakan untuk pengambilan keputusan; (3) model dapat mendapatkan pengetahuan (aturan memasukkan dan ekstraksi) untuk memperbaiki kualitas pemodelan. Yang ada neural network tinggi presisinya dalam kualitas pemodelan tetapi rendah dalam dua indikator terakhir. Neural network sebenarnya bisa dilihat sebagai kotak hitam untuk pengguna, pembatasan aplikasi membuat klasifikasi dan proses prediksi tidak dapat dipahami oleh pengguna dan langsung digunakan untuk pengambilan keputusan. Untuk data mining, tidak cukup tergantung pada model jaringan saraf menyediakan hasil karena sebelum itu pembuatan keputusan yang penting yaitu penggunaperlu memahami alasan dan pembenaran untuk pengambilan keputusan. Oleh karena itu, dalam data mining, neural network pengetahuan dasarnya harus dibentuk dalam rangka menerima domain pengetahuan dan pengetahuan neural network belajar untuk sistem dalam proses data mining. Artinya, dalam neural network di data mining, perlu menggunakan metode pengetahuan untuk mengekstraksi pengetahuan dari proses data mining dan menyadari *inoscultation* pengolahan pengetahuan dan jaringan syaraf. Selain itu, dalam sistem keputusan yang efektif dan

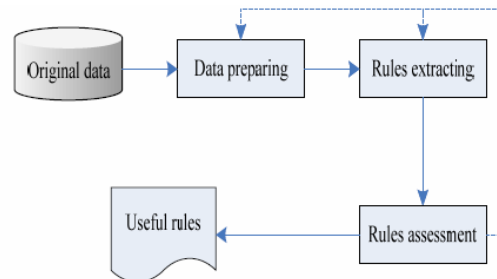
jelas.

Mekanisme juga harus didirikan untuk meningkatkan validitas dan kepraktisan dari data mining serta teknologi neural network.

- c. Perangkat muka input dan output.

Berdasarkan pertimbangan bahwa metode yang menggunakan perangkat neural network sangatlah kompleks, diperlukan perangkat muka dengan basisdata relasional, basisdata multidimensional dan *data warehouse* harus dapat disinergikan untuk kebutuhan *data mining*.

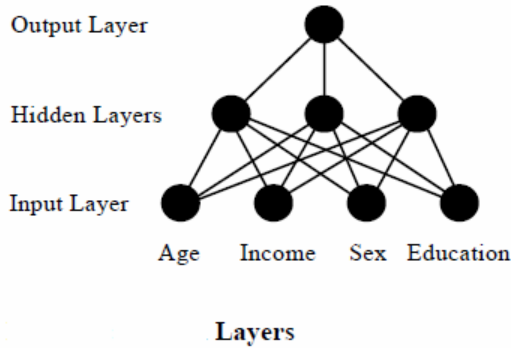
*Data mining* berdasarkan *neural network* disusun dengan menyiapkan data, *rules extracting*, *rules assessment*.



Data mining process based on neural network

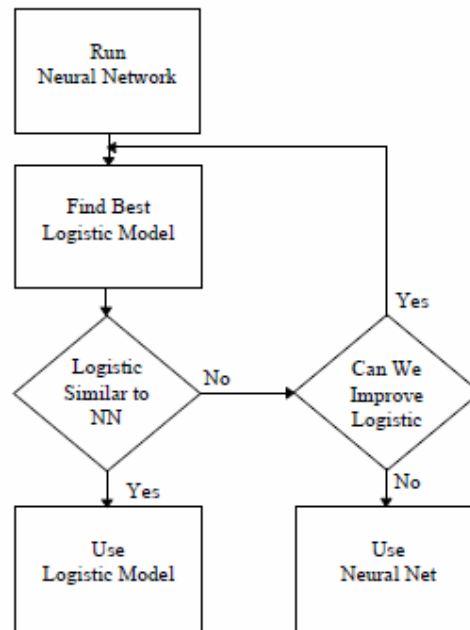
Sebagai contoh, kita mempunyai database yang terdiri dari *records* ribuan orang yang memesan brosur tentang produk terbaru kita. Database terdiri dari 5 variabel; umur, jenis kelamin, pendapatan, tingkat pendidikan, dan apakah dia memesan produk kita setelah *me-review* informasi yang kita sediakan. Kita dapat membangun neural network dengan empat node input (umur, jenis kelamin, pendapatan, tingkat pendidikan), satu node output (*behaviour* pemesanan), dan satu layer node tersembunyi. Kita dapat melatih *network* menggunakan database. Setelah itu, kita dapat menggunakannya untuk mengirimkan brosur kepada orang yang diklasifikasikan sebagai *client* oleh *neural network*.

Kekuatan neural network adalah dengan satu layer tersembunyi fungsi *transfer logistic*, *neural network* dapat mengaproksimasi tipe-tipe fungsi non linear dan *interaction terms*.



Kita menggunakan tiga langkah untuk mendekati jembatan antara *neural network* dan *logistic regression*. Pertama kita menjalankan algoritma *neural network* di data. Langkah pertama digunakan untuk menghasilkan *benchmark* solusi yang terbaik. Kedua, kita menerapkan algoritma yang sudah diseleksi di model *logistic regression* untuk menemukan sekumpulan variabel dan interaksi yang paling relevan dengan masalah. Ketika kita membandingkan *neural network* (NN) dan model *logistic*. Jika hasilnya hampir sama, maka model *logistic* adalah esensi dari relasi yang digambarkan *neural network*. Jika hasilnya berbeda, maka model *logistic* kehilangan komponen yang krusial dari perbedaan diantara *client* dan *non-client*. Sehingga kita perlu melihat lebih dekat variabel yang terdapat dalam model dan mencari variabel mana yang

hilang.



### Integrative Approach

## 6. Kesimpulan

Dari makalah ini, kita dapat melihat peranan neural network dalam data mining. Neural network merupakan topik yang relatif baru dan masih memerlukan penelitian untuk menghasilkan inovasi. Data mining sangat sesuai untuk memecahkan masalah karena akurasi yang tinggi, toleransi terhadap *noise*, independensi dari asumsi *prior*, mudah untuk dikelola. Juga *neural network* dapat dilatih di dataset yang sangat besar secara iteratif. *Data mining* menjadi lebih berguna dengan menggabungkan kekuatan *neural network* dengan perangkat statistik. Diercaya dengan gabungan ini, dapat menghasilkan sinergi yang penting. *Neural network* menyediakan benchmark model yang baik.

Dengan menggabungkan kekuatan data mining dan metode yang tradisional, kita dapat mendapatkan pengertian yang baik

terhadap lingkungan *marketing*. Ini akan mengarahkan kita kepada keputusan *marketing* yang lebih baik dan akhirnya akan membawa keuntungan yang lebih besar kepada perusahaan.

**Daftar Pustaka:**

A. Ainslie dan X.Dreze. *Data Mining: Using Neural Networks as a Benchmark for Model Building*.

Jianlin Cheng. *A Neural Network Approach to Ordinal Regression*.

Yang Gao. *Learning Classifier System Ensemble for Data Mining*. Nanjing University.

Yashpal Singh. *Neural Networks In Data Mining*. United Institute of Engineering & Technology India.

Xianjun Ni. *Research of Data Mining Based on Neural Networks*