

## DATA MINING UNTUK PREDIKSI KEBANGKRUTAN PERUSAHAAN BERDASARKAN DATA KUALITATIF

Ahmad Setiadi

AMIK BSI Karawang  
Jl. Banten No. 1 Karangpawitan, Karawang  
Telp : (0267)8454893  
E-Mail : [ahmad.ams@bsi.ac.id](mailto:ahmad.ams@bsi.ac.id)

**Abstract** - *Qualitative data are subjective and usually more difficult to measure. The objective of this research is predicting qualitative bankruptcy using data mining with neural network model. Multilayer Perceptron (MLP) and Radial Basis Function (RBF) as neural network algorithms are chosen to compare because these algorithms are commonly used. The data taken from UCI machine learning repository with 250 records and 6 attributes and SPSS Neural Network 17.0 is used to implement it. The Confusion Matrix (CM) and ROC (Receiver Operating Characteristic) Curve method is used to measure the performance of the both algorithms. Based on the test results of the implementation, proved that MLP algorithm has higher value of accuracy than RBF algorithm. Using Confusion Matrix, MLP algorithm has higher value of accuracy with 98.7% than RBF algorithm with 97.4%. Using the ROC Curve, MLP algorithm also has higher AUC (Area Under the Curve) value with 1.000 than the RBF algorithm with 0.998. From both of the algorithms, the accuracy values are included as excellent classification, because the AUC value is in the range of 0.90 until 1.00.*

**Keywords:** *data mining, bankrupt, qualitative data*

### I. PENDAHULUAN

Persaingan di bidang ekonomi semakin ketat, seiring perubahan skala persaingan ekonomi yang tidak lagi hanya pada tingkat nasional, tetapi sudah menuju pada persaingan secara regional, bahkan global. Salah satunya ditandai dengan akan terealisasinya ASEAN *Economic Community*; satu dari tiga pilar utama ASEAN *Community* 2015.

ASEAN *Economic Community* memang menghadirkan kesempatan yang menguntungkan bagi negara-negara di kawasan Asia Tenggara untuk meningkatkan daya saing melalui liberalisasi ekonomi yang terpadu dan lebih baik, serta kerjasama yang lebih erat [1]. Namun, hadirnya era liberalisasi ekonomi tersebut menuntut tata kelola keuangan perusahaan yang lebih baik agar mampu beradaptasi dengan perubahan yang terus terjadi.

Perusahaan menghadapi tantangan besar, yaitu bagaimana menghindari perusahaan dari kesulitan keuangan (*financial distress*), agar tidak mengalami krisis keuangan (*financial crises*) yang pada akhirnya membawa perusahaan pada kebangkrutan (*bankrupt*).

Kebangkrutan adalah suatu kondisi dimana suatu perusahaan tidak mampu lagi untuk mengoperasikan perusahaan dengan baik karena kesulitan keuangan yang dialami entitas tersebut sudah sangat parah [2]. Kebangkrutan atau kegagalan keuangan biasanya terjadi ketika jumlah kewajiban melebihi nilai wajar aset atau ketika kewajiban lancar melebihi aktiva lancar [3].

Kebangkrutan atau kegagalan keuangan harus dihindari karena akan membawa dampak yang luas. Kebangkrutan yang dialami oleh suatu perusahaan

besar dapat berdampak buruk terhadap perekonomian dunia [4].

Adanya potensi kebangkrutan yang dimiliki oleh setiap perusahaan mengkhawatirkan berbagai pihak, baik pihak internal perusahaan seperti manajer dan karyawan, maupun pihak eksternal perusahaan seperti investor dan kreditor. Kebangkrutan akan menyebabkan investor kehilangan saham yang ditanamkan di perusahaan tersebut dan pihak kreditor akan mengalami kerugian karena telah meminjamkan modal yang tidak akan bisa dilunasi pihak perusahaan (tak tertagih). Karena dampak yang begitu besar itulah, maka analisis prediksi kebangkrutan sangat diperlukan dalam pengambilan keputusan investasi [5].

Prediksi kebangkrutan sudah menjadi fokus pada analisis bisnis karena pentingnya keakuratan dan kecepatan waktu dalam pengambilan keputusan bisnis yang strategis. Meskipun keakuratan dari model prediksi adalah kriteria yang penting, kemampuan pemahaman dan kemampuan penerapan dari model juga penting. Akurasi prediksi kebangkrutan menjadi masalah kritis bagi pemilik modal, kreditor, para pembuat keputusan dan manajer-manajer bisnis [6].

Prediksi kebangkrutan dirasakan penting dalam meningkatkan tata kelola perusahaan. Ekonomi yang mengglobal menjadikan perusahaan harus semakin serius mengelola resiko yang mungkin terjadi dalam pengelolaan kewajiban perusahaan untuk meminimalkan resiko kredit. Banyak model yang digunakan untuk memprediksi kebangkrutan. Masing-masing model memiliki kelebihan dan kekurangan, dan memilih satu diantara metode tersebut untuk diterapkan secara empiris tidaklah mudah [7].

Prediksi kebangkrutan dapat digunakan untuk membantu berbagai pihak yang akan terkena dampaknya. Metode prediksi berkembang secara konstan dan *artificial neural network* sekarang mendapatkan tempat khusus diantara metode-metode tersebut. *Artificial neural network* ini berada diantara sistem dinamis yang mentransmisikan pengetahuan atau aturan-aturan di balik data menjadi struktur-struktur jaringan melalui pemrosesan data eksperimental [8].

Model *neural network* terbukti lebih akurat, baik dalam pembentukan model (*training process*) maupun dalam hasil pengujian (*test result*). Dalam beberapa tahun terakhir, *neural network* telah muncul sebagai sebuah teknologi yang dapat mengidentifikasi dan memodelkan pola-pola data; suatu tugas yang mungkin tidak mudah dilakukan jika menggunakan metode statistik tradisional [8].

Prediksi kebangkrutan dapat dilakukan menggunakan dua jenis parameter, yaitu parameter kuantitatif dan parameter kualitatif. Banyak metode-metode prediksi kuantitatif yang tersedia. Metode-metode ini telah meningkatkan akurasi prediksi kebangkrutan secara kuantitatif. Banyak organisasi-organisasi yang telah menerapkan metodologi ini untuk memprediksi kebangkrutan. Terlepas dari telah terbuktinya metodologi-metodologi ini untuk memprediksi kebangkrutan secara kuantitatif, pada kenyataannya banyak organisasi yang cenderung seringkali gagal [9].

Berdasarkan uraian di atas, dalam penelitian ini, penulis ingin mengetahui apakah *data mining*, khususnya model *neural network* dapat digunakan untuk memprediksi kebangkrutan perusahaan berdasarkan data kualitatif. penelitian dilakukan dengan mengolah data yang bersumber dari alamat web <http://archive.ics.uci.edu/ml/> dengan menggunakan perangkat lunak SPSS *Neural Network* 17.0.

Penelitian dilakukan untuk mengetahui tingkat akurasi yang dihasilkan oleh dua algoritma model *neural network*, yaitu *Multilayer Perceptron (MLP)* dan *Radial Basis Function (RBF)*. Kedua algoritma tersebut dipilih karena MLP dan RBF merupakan algoritma yang paling umum digunakan [10]. Tingkat akurasi yang dihasilkan algoritma MLP dan RBF ini kemudian akan dibandingkan menggunakan dua metode evaluasi, yaitu *Confusion Matrix (CM)* dan kurva ROC (*Receiver Operating Characteristic*).

Dari penelitian ini diharapkan dapat diketahui apakah *data mining* dapat digunakan untuk memprediksi kebangkrutan perusahaan berdasarkan data kualitatif melalui pembuktian tingkat akurasi prediksi yang dicapai. Melalui penelitian ini juga diharapkan akan dapat diketahui, algoritma *neural network* yang memiliki tingkat akurasi yang lebih tinggi, sehingga dapat digunakan untuk memprediksi kebangkrutan perusahaan.

## II. LANDASAN TEORI

### 2.1. Data Mining

*Data Mining* [11] didefinisikan sebagai proses penemuan pola dalam data. Gartner Group [12] menyebutkan bahwa *data mining* adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan teknik matematika.

Menurut Daryl Pregibons [13] “*Data mining* adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang *database*”. Nama *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *database* yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai [14].

*Data mining* merupakan pemecahan masalah dengan menganalisa data yang sudah ada dalam *database*. *Data mining* didefinisikan sebagai proses menemukan pola dalam data. Proses ini harus otomatis atau biasa juga secara semi otomatis. Pola yang dihasilkan harus berarti, artinya pola tersebut memberikan beberapa keuntungan. Dalam *data mining*, data disimpan secara elektronik dan pencarian datanya dilakukan secara otomatis atau ditambahkan oleh komputer. Para ekonom, ahli statistik, peramal, dan teknisi-teknisi komunikasi telah lama bekerja dengan ide bahwa pola-pola dalam sebuah data dapat dicari secara otomatis, diidentifikasi, divalidasi, dan digunakan untuk membuat sebuah prediksi [11].

### 2.2. Algoritma Klasifikasi Data Mining

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [15].

Klasifikasi data terdiri dari dua langkah proses, yaitu [15]:

1. Proses *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi.
2. Proses klasifikasi, dimana *data test* digunakan untuk memperkirakan akurasi dari *rule* klasifikasi.

Proses klasifikasi didasarkan pada empat komponen [13]:

- a. *Kelas*  
Variabel dependen yang berupa kategorikal yang merepresentasikan “label” yang terdapat pada objek.
- b. *Predictor*  
Variabel independen yang direpresentasikan oleh karakteristik (atribut) data.
- c. *Training dataset*  
Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk

menentukan kelas yang cocok berdasarkan *predictor*.

d. *Testing dataset*

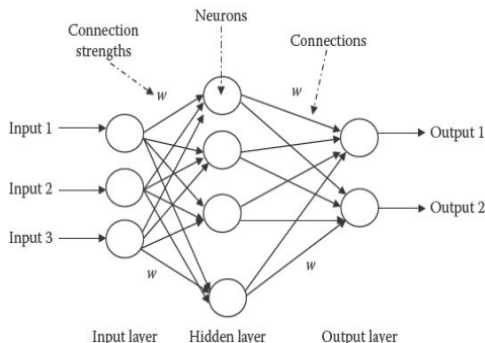
Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Algoritma klasifikasi yang paling terkenal dan banyak digunakan secara luas terdiri dari [13]:

1. *Decision/classification trees*
2. *Bayesian classifiers/ Naïve Bayes classifiers*
3. *Neural networks*
4. Analisa Statistik
5. Algoritma Genetika
6. *Rough sets*
7. *K-nearest neighbor*
8. Metode *Rule Based*
9. *Memory based reasonin*
10. *Support vector machines (SVM)*

2.3. *Neural Network*

*Neural network* berusaha untuk meniru struktur dan cara kerja otak manusia sehingga mampu menggantikan beberapa pekerjaan manusia. Pekerjaan seperti mengenali pola (*pattern recognition*), prediksi, klasifikasi, pendekatan fungsi optimasi adalah pekerjaan-pekerjaan yang diharapkan bisa diselesaikan dengan *neural network* [16].



Gambar 1. Model *Neural Network* [17]

*Neural network* terdiri dari sejumlah input. Informasi ini diberikan sebagai masukan melalui input koneksi, masing-masing koneksi memiliki beberapa bobot yang berhubungan dengan input tersebut. Masukan tambahan yang dikenal sebagai bias, diberikan ke *artificial neuron*. *Neuron* juga terdiri dari satu *output*. *Outputnya* adalah terbentuk dari pengolahan berbagai input oleh *neuron* [18].

Setiap simpul dalam *neural network* merupakan sebuah unit pemrosesan. Tiap simpul memiliki beberapa masukan dan sebuah keluaran. Setiap simpul mengkombinasikan beberapa nilai masukan, melakukan kalkulasi, dan membangkitkan nilai keluaran (aktifasi). Dalam setiap simpul terdapat dua fungsi, yaitu fungsi untuk mengkombinasikan masukan dan fungsi aktifasi untuk menghitung keluaran [17].

2.4. *Multilayer Perceptron*

*Multilayer Perceptron (MLP)* disebut juga *multilayer feedforward neural network* merupakan algoritma yang paling luas digunakan. Menurut Wong, Bodnovich dan Selvi (1997), sekitar 95% aplikasi bisnis yang menggunakan *neural network*, memakai algoritma ini [19]. Salah satu kelebihan *neural network* adalah cukup baik dalam menangani data yang mengandung *noise* [12].

MLP terdiri dari tiga *layer*, yaitu [19]:

1. Lapisan Input (*Input layer*)  
*Input layer* untuk menerima nilai masukan dari tiap *record* pada data. Jumlah simpul input sama dengan jumlah variabel *prediktor*.
2. Lapisan Tersembunyi (*Hidden layer*)  
*Hidden layer* mentransformasikan nilai input di dalam *network*. Tiap simpul pada *hidden layer* terhubung dengan simpul-simpul pada *hidden layer* sebelumnya atau dari simpul-simpul pada *input layer* dan ke simpul-simpul pada *hidden layer* berikutnya atau ke simpul-simpul pada *output layer*. Jumlah *hidden layer* bisa berapa saja.
3. *Output layer*  
Garis yang terhubung dengan *Output layer* berasal dari *hidden layer* atau *input layer* dan mengembalikan nilai keluaran yang bersesuaian dengan variabel prediksi. Keluaran dari *output layer* biasanya merupakan nilai *floating* antara 0 sampai 1 [17].

2.5. *Radial Basis Function*

*Radial Basis Function (RBF)* memiliki kemampuan yang baik dalam pemodelan data non linier dan model dapat dibentuk dalam satu tahap, berbeda dengan MLP yang harus beberapa kali perulangan, sehingga menghasilkan *output* aplikasi yang lebih cepat. RBF sangat berguna untuk memecahkan masalah dimana data input tidak murni disebabkan *noise*. Fungsi transformasi yang digunakan didasarkan pada distribusi *Gaussian*. Jika kesalahan pada jaringan diminimalisasikan dengan tepat akan menghasilkan *output* berupa suatu penjumlahan, yang menunjukkan kemungkinan dari *output*. Namun demikian, RBF memiliki keterbatasan, yaitu lebih sensitif secara dimensi dan memiliki sedikit kesulitan jika jumlah unit besar [20].

Jaringan RBF memiliki tiga lapisan:

1. Lapisan Input (*Input Layer*)  
Ada 1 *neuron* dalam lapisan input pada setiap variabel *predictor*. *Neuron* masukan atau proses sebelum lapisan masukan menstandarkan jarak dari nilai dengan cara menambahkan median dan pembagian melalui jarak *interquartile*. *Neuron* masukan kemudian memasukkan nilai pada setiap *neuron* yang ada pada lapisan tersembunyi.
2. Lapisan Tersembunyi (*Hidden Layer*)  
Lapisan ini memiliki beberapa variabel *neuron* (beberapa diantaranya ditentukan melalui proses

pengujian). Setiap *neuron* terdiri dari sebuah RBF yang terpusat pada satu titik yang memiliki banyak dimensi sebagaimana variabel *predictor*. Jarak dari RBF mungkin berbeda pada setiap dimensinya. Pusat dan jarak dibedakan melalui proses pengujian. Ketika ditunjukkan dengan vektor  $x$  dari nilai masukan ke lapisan input sebuah *neuron* tersembunyi menghitung jarak *euclidean* sebuah *test case* dari titik pusat *neuron* lalu mengaplikasikan ke RBF ke fungsi *kernel* RBF untuk jarak tersebut menggunakan nilai jarak. Nilai hasil dilewati melalui lapisan penjumlahan.

3. Lapisan Penjumlahan (*Summation Layer*)  
Nilai yang datang dari *neuron* dalam lapisan tersembunyi digandakan melalui penjumlahan (penggabungan) bobot dengan *neuron* dan menuju penjumlahan yang menambahkan nilai bobot dan menunjukkan penjumlahannya sebagai hasil (*output*) dari *network*.

## 2.6. Evaluasi dan Validasi Klasifikasi Data Mining

Evaluasi dan validasi hasil klasifikasi dengan *data mining* pada penelitian ini menggunakan metode *Confusion Matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

### 1. Confusion Matrix

Metode ini hanya menggunakan tabel matriks seperti padaterlihat pada Tabel 1. Jika *dataset* hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [21]. Evaluasi dengan confusion matrix menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [15].

Tabel 1. Model *Confusion Matrix* [15].

<i>Correct Classification</i>	<i>Classified as</i>	
	+	-
+	<i>True positives</i>	<i>False negatives</i>
-	<i>False positives</i>	<i>True negatives</i>

*True Positive* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah *record negative* yang diklasifikasikan sebagai positif, *false negative* adalah jumlah *record* positif yang diklasifikasikan sebagai negative, *true negative* adalah jumlah *record negative* yang diklasifikasikan sebagai negative.

### 2. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *Confusion Matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* sebagai garis vertical [19]. *Area Under Curve* (AUC) dihitung

untuk mengukur perbedaan performansi metode yang digunakan.

Pada *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok, yaitu [13]:

- a. 0.90 sampai 1.00 : Klasifikasi sangat baik (*Excellent Classification*)
- b. 0.80 sampai 0.90 : Klasifikasi baik (*Good Classification*)
- c. 0.70 sampai 0.80 : Klasifikasi cukup (*Fair Classification*)
- d. 0.60 sampai 0.70 : Klasifikasi buruk (*Poor Classification*)
- e. 0.50 sampai 0.60 : klasifikasi salah (*Failure Classification*)

## 2.7. Prediksi Kebangkrutan

Pada penelitian awal mengenai teori kebangkrutan, bangkrut didefinisikan sebagai kegagalan suatu perusahaan mendapatkan laba yang meningkatkan kemungkinan perusahaan tidak mampu membayar pinjaman pokok dan bunganya [22].

Definisi lain dari bangkrut adalah suatu kondisi dimana arus kas perusahaan lebih sedikit dibandingkan total biaya bunga pinjaman jangka panjang [23]. Kebangkrutan terjadi ketika suatu perusahaan menderita banyak kerugian atau perusahaan gagal untuk membayar hutangnya. Kegagalan dalam pembayaran hutang terjadi ketika perbandingan asset perusahaan terhapus oleh hutang, dan perusahaan gagal untuk membayar hutangnya.[24].

Kebangkrutan adalah pernyataan resmi mengenai ketidakmampuan atau ketidakberdayaan dari seseorang atau organisasi untuk membayar pinjaman kepada kreditur-kreditur mereka. Prediksi kebangkrutan sangat penting karena menyediakan dua tujuan utama dibawah hukum kebangkrutan. Pertama, hukum kebangkrutan memberikan kreditur beberapa pembayaran pada hutang mereka jika debitur (seseorang yang berhutang pinjaman) memiliki kemampuan untuk membayarnya. Kedua, hukum kebangkrutan memberi debitur keputusan pembebasan, dengan membatalkan hutang mereka. Jika sebuah organisasi tidak menginginkan berada pada kondisi yang tidak diinginkan, maka dapat menerapkan teknik prediksi kebangkrutan [8].

Prediksi kebangkrutan adalah suatu metode yang dapat digunakan untuk mengetahui keuntungan yang tepat untuk peluang investasi dan pengalokasian sumberdaya yang lebih baik. Prediksi meramalkan kemungkinan kejadian di masa yang akan datang didasarkan pada informasi sekarang dan masa lalu. Kemudian menyediakan perusahaan dengan peringatan yang perlu untuk mewaspadai kebangkrutan dan mengambil tindakan yang dianggap tepat untuk mengatasi hal tersebut. Disamping itu, hal ini membuat investor dan kreditur mengetahui kesempatan investasi yang lebih menguntungkan dan menyebabkan penginvestasian sumber daya mereka pada peluang yang tepat. Oleh karenanya, prediksi

kebangkrutan perusahaan selalu menjadi salah satu hal yang penting untuk investor, kreditur dan pemerintah [25].

Teknik-teknik yang dikembangkan untuk meningkatkan prediksi kebangkrutan dikalsifikasikan menjadi tiga kelompok, yaitu model statistic, kecerdasan tiruan, dan model teoritis [8].

Kecerdasan tiruan melalui klasifikasi adalah suatu metode untuk memprediksi nilai kategori dari suatu atribut target, tidak seperti model statistik yang atributnya harus berbentuk numerik. Dimulai dari suatu kumpulan data masa lalu yang diobservasi dimana kelasnya diketahui, model klasifikasi digunakan untuk membentuk kumpulan model peraturan (algoritma) yang nantinya dapat digunakan untuk memprediksi data di masa yang akan datang [19].

Prediksi kebangkrutan adalah masalah dalam pengklasifikasian kedalam dua kelas : Bangkrut atau tidak bangkrut (sehat). Lebih dari beberapa tahun lalu krisis keuangan diamati dalam beberapa sektor, termasuk perbankan. Kebangkrutan dapat berakibat pada semua wilayah, dimana yang terkena dampaknya adalah para kreditur, auditor, pemegang saham dan manajemen senior. Oleh karenanya, semua tertarik dalam prediksi kebangkrutan [8].

## 2.8. Data Kualitatif

Menurut sifatnya data dapat terbagi menjadi dua jenis, yaitu data kualitatif (non metrik) dan data kuantitatif (metrik). Kemudian jenis data kualitatif terbagi lagi menjadi dua jenis, yaitu data nominal dan data ordinal. Begitu pula dengan jenis data kuantitatif terbagi menjadi dua jenis, yaitu data interval dan data rasio [26].

Data kualitatif secara sederhana dapat disebut data yang bukan berupa angka. Ciri utama data

kualitatif didapat dengan cara menghitung, sehingga tidak memiliki nilai desimal. Selain itu data kualitatif memiliki ciri-ciri tidak bisa dilakukan operasi matematika, seperti penambahan, pengurangan, perkalian dan pembagian. Contoh data kualitatif adalah data gender, data golongan darah, data tempat tinggal atau data jenis pekerjaan. Agar dapat dilakukan proses pada data kualitatif atau non metrik, data tersebut harus diubah ke dalam bentuk angka, proses ini dinamakan kategorisasi [16].

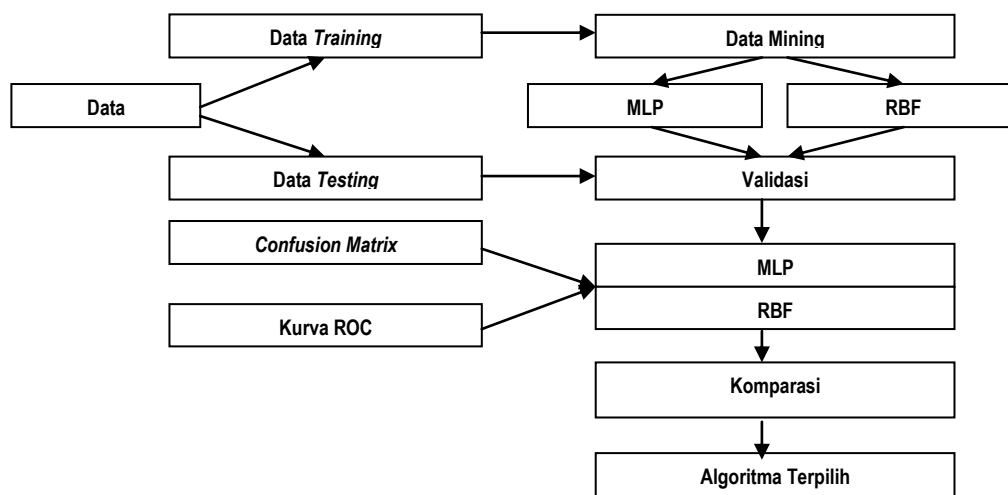
Data Nominal adalah jenis data kualitatif yang digunakan untuk mengidentifikasi, mengklasifikasi, atau membedakan objek. Data nominal merupakan data yang paling rendah dalam level pengukuran data. Semua data memiliki posisi setara dalam arti tidak ada data yang memiliki tingkat yang lebih atau kurang dibandingkan dengan data yang lain. Jenis data nominal ini tidak memiliki jarak, urutan dan titik origin [26].

Data Ordinal adalah jenis data kualitatif namun memiliki level lebih tinggi dari data nominal. Data ordinal memiliki karakteristik nominal tapi terdapat perbedaan derajat, urutan, atau peringkat dalam objek tersebut (posisi data tidak setara) [26].

## III. PEMBAHASAN

### 3.1 Kerangka Pemikiran Penelitian

Jenis penelitian yang dilakukan dalam penelitian ini adalah jenis penelitian eksperimen komparatif. eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik [27]. Penelitian ini dilandasi kerangka pemikiran berikut:



Gambar 2. Kerangka Pemikiran Penelitian

### 3.1. Pengumpulan Data

Kebanyakan kumpulan data yang dibutuhkan untuk *data mining* tidak tersedia bagi umum untuk digunakan. Namun, sejumlah ‘perpustakaan’

kumpulan data telah tersedia untuk diunduh dari *World WideWeb* dan gratis untuk setiap orang. Perpustakaan yang paling terbaik adalah repositori

kumpulan data yang dimaintain oleh University of California at Irving, yang umumnya dikenal sebagai 'UCI Reppository' [21].

Data yang digunakan dalam penelitian ini bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>. parameter data dikumpulkan berdasarkan analisis ahli. Sesuai dengan analisis para ahli, enam parameter dijadikan sebagai

parameter-parameter paling penting untuk menyebabkan aturan-aturan seperti Resiko Industri atau *Industry Risk* (IR). Resiko manajemen atau *Management Risk* (MR), Fleksibilitas keuangan atau *Financial flexibility* (FF), Kredibilitas atau *Credibility* (CR), kemampuan kompetisi atau *Competitiveness* (CO) dan Resiko Operasi atau *Operation Risk* (OR) [9].

Tabel 2 Atribut Data yang digunakan dalam Penelitian

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	IR	String	1	0		None	None	5	Center	Nominal
2	MR	String	1	0		None	None	5	Center	Nominal
3	FF	String	1	0		None	None	5	Center	Nominal
4	CR	String	1	0		None	None	5	Center	Nominal
5	CO	String	1	0		None	None	5	Center	Nominal
6	OP	String	1	0		None	None	5	Center	Nominal
7	Class	String	2	0		None	None	6	Center	Nominal

Record yang digunakan berjumlah 250. Masing-masing record terdiri dari 6 atribut yang merupakan parameter kuantitatif untuk prediksi kebangkrutan. Setiap atribut berisi nilai parameter prediktor berupa P untuk *Positive* (baik), A untuk *Average* (sedang), dan

N untuk *Negative* (buruk). Prediksi kebangkrutan diklasifikasikan menjadi dua kelas, yaitu B, yaitu *Bankrupt* (Bangkrut) dan NB, yaitu *Non-Bankrupt* (Sehat). Distribusi data terdiri dari 143 record untuk *Non-Bankrupt* dan 107 untuk *Bankrupt*.

Tabel 3. Record yang digunakan dalam Penelitian

	IR	MR	FF	CR	CO	OP	Class
1	P	P	A	A	A	P	NB
2	N	N	A	A	A	N	NB
3	A	A	A	A	A	A	NB
4	P	P	P	P	P	P	NB
5	N	N	P	P	P	N	NB
6	A	A	P	P	P	A	NB
7	P	P	A	P	P	P	NB
8	P	P	P	A	A	P	NB
9	P	P	A	P	A	P	NB
10	P	P	A	A	P	P	NB
11	P	P	P	P	A	P	NB
12	P	P	P	A	P	P	NB
13	N	N	A	P	P	N	NB
14	N	N	P	A	A	N	NB
15	N	N	A	P	A	N	NB
16	N	N	A	P	A	N	NB

### 3.2. Pengumpulan Data

Untuk menyusun data, ditentukan inisiasi pembangkit aktif (*active generator initialization*), yaitu nilai awal (*starting point*) berupa nilai tetap (*fixed value*) : 2000000. Data yang ada dipilah terdiri dari 70% data

digunakan untuk pembentukan model (*Data Training* atau *Training Sample*) dan 30% digunakan untuk menguji model (*Data Testing* atau *Testing Sample*).

Tabel 4. Pembagian *Data Training* dan *Testing*

	MLP		RBF	
	Jumlah	Percent	Jumlah	Percent
<i>Sample Training</i>	174	69.6%	174	69.6%
<i>Testing</i>	76	30.4%	76	30.4%
<i>Valid</i>	250	100%	250	100%
<i>Exclude</i>	0		0	
<b>Total</b>	<b>250</b>		<b>250</b>	

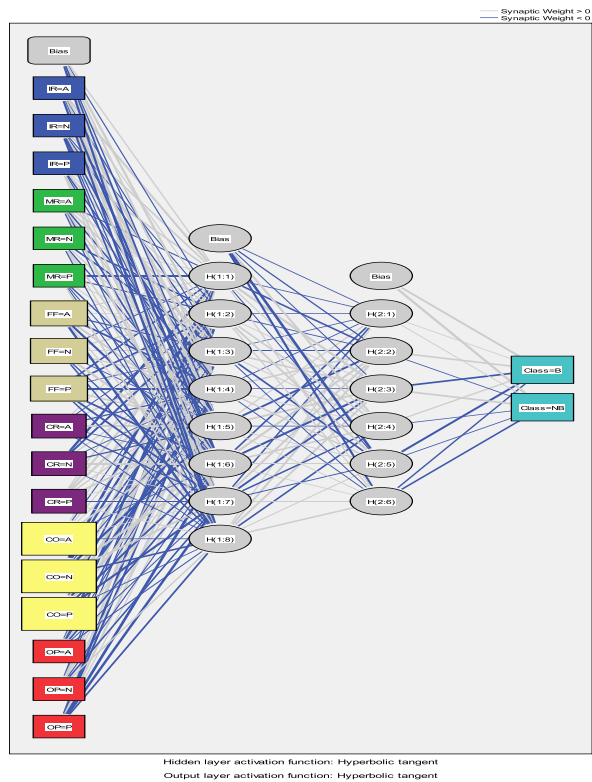
### 3.3. Penerapan dan Evaluasi Algoritma Neural Network

Data yang telah dipilah menjadi *data training* dan

*data testing* kemudian diterapkan kedalam algoritma *neural network*, yaitu MLP dan RBF.

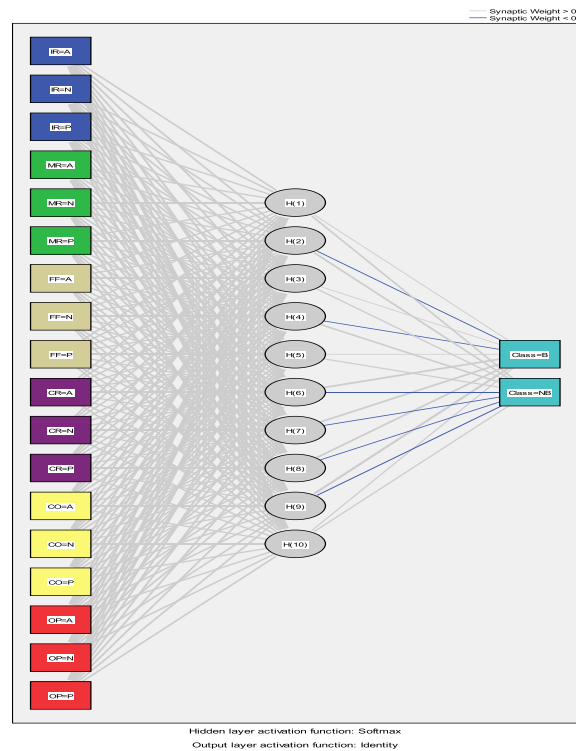


Neural network yang terbentuk dengan MLP adalah:



Gambar 3. Neural Network MLP

Neural network yang terbentuk dengan RBF adalah:



Gambar 4. Neural Network RBF

Hasil penerapan kedua algoritma tersebut dievaluasi dengan metode *Confusion Matrix* dengan hasil:

Tabel 5. *Confusion Matrix*

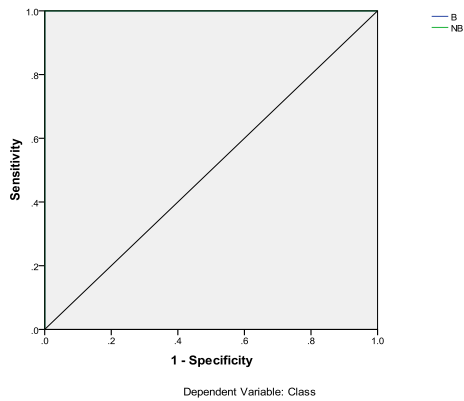
Sample	Observed	Predicted MLP			Predicted RBF		
		B	NB	Percent Correct	B	NB	Percent Correct
Training	B	79	0	100.0%	76	3	96.2%
	NB	0	95	100.0%	0	95	100.0%
	Overall Percent	45.4%	54.6%	100.0%	43.7%	56.3%	98.3%
Testing	B	27	1	96.4%	26	2	92.9%
	NB	0	48	100.0%	0	48	100.0%
	Overall Percent	35.5%	64.5%	98.7%	34.2%	65.8%	97.4%

Dependent Variable: Class

Penerapan algoritma MLP dan RBF juga dievaluasi menggunakan kurva ROC. Berdasarkan evaluasi

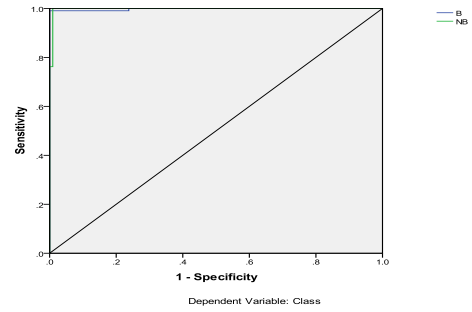
yang dilakukan, untuk algoritma MLP dihasilkan nilai AUC (*Area Under the Curve*) 1.000 dan RBF 0.998.

Kurva ROC MLP adalah :



Gambar 5. Kurva ROC MLP

Sedangkan Kurva ROC untuk RBF adalah:



Gambar 6. Kurva ROC RBF

### 3.4. Komparasi Algoritma *Neural Network*

Untuk mengetahui algoritma mana yang memiliki akurasi yang lebih tinggi dalam memprediksi kebangkrutan, digunakan metode *Confusion Matrix*

dan Kurva ROC dengan hasil komparasi yang terlihat pada tabel 6:

Tabel 6. Hasil Komparasi

	<b>MLP</b>	<b>RBF</b>
<i>Confusion Matrix</i>	98.7%	97.4%
Kurva ROC (AUC)	1.000	0.998

Perbedaan hasil prediksi dari MLP dan RBF dapat terlihat pada gambar berikut:

Tabel 7. Komparasi Variabel Dependen Class dengan Hasil prediksi MLP dan RBF

	IR	MR	FF	CR	CO	OP	Class	MLP_PredictedValue	MLP_PseudoProbability_1	MLP_PseudoProbability_2	RBF_PredictedValue	RBF_PseudoProbability_1_A	RBF_PseudoProbability_2_A
135	N	P	A	A	A	N	NB	NB	0.026	0.971	NB	0.048	0.952
136	P	N	A	A	A	N	NB	NB	0.002	0.969	NB	0.050	0.950
137	P	N	P	P	P	P	NB	NB	0.033	0.974	NB	0.047	0.953
138	N	P	P	P	P	N	NB	NB	-0.023	0.973	NB	0.047	0.953
139	P	N	P	P	P	N	NB	NB	-0.014	0.972	NB	0.048	0.952
140	N	N	A	P	P	P	NB	NB	0.032	0.973	NB	0.047	0.953
141	P	N	P	A	A	P	NB	NB	0.019	0.974	NB	0.047	0.953
142	N	P	A	P	A	P	NB	NB	-0.019	0.971	NB	0.047	0.953
143	N	P	A	A	P	N	NB	NB	0.013	0.973	NB	0.048	0.952
144	A	N	N	N	N	A	B	B	0.986	-0.050	B	1.036	-0.036
145	P	N	N	N	N	N	B	B	0.989	-0.020	B	1.093	-0.093
146	N	P	N	N	N	N	B	B	0.989	-0.014	B	1.058	-0.058
147	A	P	N	A	N	N	B	B	0.982	0.079	NB	0.354	0.646
148	N	N	N	N	N	N	B	B	0.989	-0.023	B	1.111	-0.111
149	N	N	N	A	N	A	B	B	0.985	-0.042	B	0.805	0.195

Terlihat pada gambar di atas, pada record 147, dimana Class bernilai B, hasil prediksi dengan MLP adalah B (tepat) dan hasil prediksi dengan algoritma RBF adalah NB (tidak tepat). Berdasarkan pengelompokan nilai AUC [13], maka dapat disimpulkan bahwa MLP dan RBF termasuk dalam klasifikasi sangat baik (*excellent classification*) karena memiliki nilai AUC antara 0.90 sampai 1.00.

### IV. KESIMPULAN

Penelitian ini bertujuan untuk mengetahui apakah *data mining* dapat digunakan untuk memprediksi kebangkrutan perusahaan berdasarkan data kualitatif. Model yang digunakan adalah *neural network* dengan algoritma MLP dan RBF. Hasil penerapan menggunakan algoritma MLP dan RBF diukur tingkat akurasinya menggunakan metode pengujian *Confusion Matrix* dan Kurva ROC.



Berdasarkan hasil pengukuran tingkat akurasi kedua algoritma tersebut, diketahui bahwa algoritma MLP memiliki tingkat akurasi yang lebih tinggi dibandingkan RBF. Penelitian ini juga menyimpulkan bahwa hasil pengukuran menggunakan metode ROC Curve terhadap penerapan algoritma MLP dan RBF menghasilkan nilai AUC yang termasuk dalam klasifikasi sangat baik (*excellent classification*). Dengan demikian, dapat disimpulkan bahwa *data mining* dengan algoritma *neural network* MLP dan RBF dapat digunakan untuk prediksi kebangkrutan perusahaan, dan algoritma MLP memiliki nilai akurasi yang lebih tinggi dibandingkan RBF.

#### DAFTAR REFERENSI

- [1] Michael G. Plummer, Chia Siow Yue. *Realizing The ASEAN Economic Community; A Comprehensive Assessment*. Singapore: ISEAS Publishing. 2009
- [2] Ni Made Evi Dwi Prihanthini1 . Maria M. Ratna Sari. Prediksi Kebangkrutan dengan Model Grover, Altman Z-Score, Springate dan Zmijewski pada Perusahaan *Food and Beverage* di Bursa Efek Indonesia.. ISSN: 2302-8556 E-Jurnal Akuntansi Universitas Udayana 5.2 (2013): 417-435
- [3] Elmabrok, Ali Abusalah., Mohammed and Ng Kim-Soon. *Using Altman's Model and Current Ratio to Assess the Financial Status of Companies Quoted In the Malaysian Stock Exchange*. International Journal of Scientific and Research Publications, 2(7). Faculty of Technology Management, Business and Entrepreneurship, Universiti Tun Hussein Onn Malaysia. 2012.
- [4] June Li. *Prediction of Corporate Bankruptcy from 2008 Through 201*. Journal of Accounting and Finance 12(1). University of Wisconsin, River Falls. 2012.
- [5] Etta Citrawati Yuliasary Made Gede Wirakusuma Analisis Financial Distress dengan Metode Z-Score Altman, Springate, Zmijewski. ISSN: 2302-8556 E-Jurnal Akuntansi Universitas Udayana 6.3 (2014):379-389
- [6] David L. Olson, Dursun Delen, Yanyan Meng. *Comparative analysis of data mining methods for bankruptcy prediction*. Journal Decision Support Systems and Electronic Commerce. 52 (2012) 464–473 United States : Elsevier. 2012.
- [7] M. Adnan Aziz and Humayon A. Dar. Corporate Governance. Vol. 6 NO. 1. 2006, pp. 18-33, Emerald Group Publishing Limited, ISSN 1472-0701 DOI 10.1108/14720700610649436
- [8] Alireza Mehrazin, Mohammad Taghipour, Omid froutan, Bashir Ghabdian1 & Hamid Soleimani. *Radial Basis Function in Artificial Neural Network for Prediction of Bankruptcy*. International Business Research; Vol. 6, No. 8; 2013. ISSN 1913-9004 E-ISSN 1913-9012 . Published by Canadian Center of Science and Education.
- [9] Martin, T. Miranda Lakshmi, V. Prasanna Venkatesan. *An Analysis on Qualitative Bankruptcy Prediction Rules using Ant-Miner. I.J. Intelligent Systems and Applications*, 2014, 01, 36-44 Published Online December 2013 in MECS (<http://www.mecs-press.org/>) DOI: 10.5815/ijisa.2014.01.05
- [10] Yilmaz, Isik, Nazn Yalcin Erik, & Oguz Kaynar, 2010, *Academic Journals: Different types of learning algorithms of artificial neural network (ANN) models for prediction of gross calorific value (GCV) of coals*, ISSN 1992-2248(c)2010 Academic Journals
- [11] Witten, I. H., Frank, E., & Hall, M. A. *Data Mining: Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher. 2011.
- [12] Larose, D. T. *Discovering Knowledge in Data*. New Jersey: John Wiley & Sons, Inc. 2005.
- [13] Gorunescu, Florin. *Data Mining Concepts, Models and Technique*. Berlin Heidelberg: Springer Verlag. 2011.
- [14] Sumathi, & S., Sivanandam, S.N.. *Introduction to Data Mining and its Applications*. Berlin Heidelberg New York: Springer. 2006.
- [15] Han, J., & Kamber, M. *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. 2006.
- [16] Santoso, S. *Statistik Multivariat*. Jakarta: Elex Media Komputindo. 2010.
- [17] Kusriani, and Luthfi, E. T. "Algoritma Data Mining". Yogyakarta: Andi Publishing. 2009.
- [18] Shukla, A., Tiwari, R., & Kala, R. *Real Life Applications of Soft Computing*. United States of America on: Taylor and Francis Group, LLC. 2010.
- [19] Carlo Vercellis. *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Wiley & Sons Ltd. 2009.
- [20] Venkatesan, P. & S. Anitha. *Application of a radial Basis Function Neural Network for Diagnosis of Diabetes Mellitus*, Chennai, India: Current Science, Vol. 91, No. 9. 2006.
- [21] Max Bramer. *Principle of Data Mining*. London : Springer-Verlag London Limited. 2007.

- [22] Gordon, M. J. (1971). Towards a Theory of Financial Distress. *The Journal of Finance*, 26, 347-356.  
<http://dx.doi.org/10.1016/j.eswa.2008.03.020>
- [23] Whitaker, R. *The Early Stage of Financial Distress*. *Journal of Economics and Finance*, 23(2), 123-133.  
<http://dx.doi.org/10.1007/BF02745946>. 1999.
- [24] Chen, W. S., & Du, Y. K. *Using neural networks and data mining techniques for the financial distress prediction model*. *Expert systems with applications*, 36, 4075-4086.  
<http://dx.doi.org/10.1016/j.eswa.2008.03.020>. 2009.
- [25] Jones, F. J. *Current Techniques in Bankruptcy Prediction*. *Journal of Accounting Literature*, 6, 131-164. 1987.
- [26] Hidayat, T., dan Istiadah, N. *Panduan Lengkap Menguasai SPSS 19 untuk Mengolah Data Statistik Penelitian*. Jakarta: Media Kita. 2011.
- [27] C.R.Kothari. *Research Methodology Methods and Techniques*. India: New Age International Limited. 2004.

#### **Biodata Penulis**

Ahmad Setiadi, memperoleh gelar Sarjana Komputer (S. Kom), jurusan Manajemen Informatika dari STMIK Budi Luhur Jakarta. Memperoleh gelar Magister Komputer (M.Kom) pada Program Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri, Jakarta