

PREDIKSI KEBANGKRUTAN PERUSAHAAN MENGGUNAKAN ALGORITMA C4.5 BERBASIS *FORWARD SELECTION*

Hamsir Saleh

Hamsir.saleh@gmail.com

Fakultas Ilmu Komputer Universitas Ichsan Gorontalo

Abstrak

Memprediksi kebangkrutan perusahaan adalah upaya yang penting dalam mengatasi masalah manajemen perusahaan dengan tujuan utamanya adalah mengoptimalkan pengelolaan fitur yang berpengaruh dalam memprediksi kebangkrutan perusahaan. Masalah mendasar dalam *machine learning* adalah proses optimasi keputusan untuk mendapatkan fungsi kombinasi yang optimal. *Forward selection* adalah pendekatan *wrapper* yang sering digunakan dalam seleksi fitur otomatis, *forward selection* mampu menghapus fitur yang tidak relevan, mengembangkan dan menambah kualitas data, serta meningkatkan performa dan akurasi model. Penelitian ini mengusulkan algoritma C4.5 berbasis *forward selection* untuk menemukan atribut yang berpengaruh dalam peningkatan akurasi prediksi kebangkrutan perusahaan. Hasil penelitian yang telah dilakukan dengan penerapan algoritma C4.5 berbasis *forward selection* menghasilkan beberapa fitur signifikan, dalam penelitian menggunakan 250 record atribut *competitiveness* dan *credibility* menjadi fitur yang signifikan dari 6 atribut yang ada. Dalam penelitian dengan 250 record algoritma C4.5 mendapatkan hasil akurasi sebesar 99.60% dan algoritma C4.5 berbasis *forward selection* dengan akurasi sebesar 99.61%.

Kata kunci : Prediksi, kebangkrutan, C4.5, Seleksi Fitur, *forward selection*

1. Pendahuluan

Memprediksi kebangkrutan perusahaan adalah upaya yang penting dalam mengatasi masalah manajemen perusahaan dengan tujuan utamanya adalah untuk membedakan perusahaan tersebut dalam keadaan sehat atau dalam keadaan bangkrut. Selain itu, salah pengambilan keputusan di lembaga - lembaga keuangan dapat membuat kesulitan keuangan atau tertekan dan menyebabkan banyak biaya sosial yang mempengaruhi pemilik atau pemegang saham, manajer, pemerintah. Kebangkrutan adalah peristiwa penting yang bisa menimbulkan kerugian besar bagi manajemen, pemegang saham, karyawan, pelanggan dan bangsa, maka memprediksi kebangkrutan perusahaan telah menjadi topik hangat baik untuk aplikasi industri dan penelitian akademis [1].

Dalam tahun – tahun terakhir banyak penelitian telah menunjukkan bahwa teknik *machine learning* seperti *Artificial Neural Networks* (ANN), *Decision Tree* (DT), *Case Based Reasoning* (CBR), *Support Vector Machine* (SVM) dapat digunakan sebagai metode alternatif untuk prediksi kebangkrutan perusahaan [2]. Berbeda dengan teknik statistik, teknik *machine learning* tidak mengasumsi distribusi data tertentu dan secara otomatis mengekstrak pengetahuan dari sampel pelatihan [3]. Kinerja prediksi tergantung pada rincian masalah, karakteristik struktur data yang digunakan, sejauh mana dimungkinkan untuk memisahkan kelas dengan menggunakan karakteristik tersebut, dan tujuan dari klasifikasi [4].

Penelitian tentang prediksi kebangkrutan perusahaan menggunakan teknik optimasi untuk membantu merancang beragam sistem klasifikasi yang akurat, dalam perbandingan klasifikasi individu SVM (72,45%) menunjukkan akurasi prediksi yang lebih tinggi dari NN (71,02%) dan DT (70,30%), prediksi DT ensambles (75,10%, 75,78%) lebih tinggi dari NN ensambles (73,10%, 73,97%) dan SVM ensambles (73,07%, 72,85%), dalam perbandingan antara *machine learning* dengan optimasi menunjukkan peningkatan hasil akurasi dalam prediksi CO – SVM (77,53%, 77,23%) lebih baik daripada CO – DT (76,00%, 76,20%) dan CO – NN (76,52%, 76,92%) [5].

Algoritma seleksi fitur otomatis digunakan dengan tujuan memilih subset dari fitur yang ada di dataset sehingga tercapai tingkat akurasi maksimal dalam klasifikasi. Dua pendekatan umum untuk seleksi fitur otomatis adalah metode filter dan metode *wrapper*, metode filter biasanya lebih cepat dalam proses seleksi namun pendekatan *wrapper* memberikan kinerja yang lebih tinggi [6]. Ada dua pendekatan dalam seleksi fitur yaitu *forward selection* dan *backward elimination*, *forward selection* dilakukan dengan cara memasukkan variabel independen yang signifikan masuk ke dalam model satu

persatu hingga tidak ada lagi variabel independen yang signifikan untuk dimasukkan ke dalam model [7].

Terdapat beberapa aspek yang dapat memprediksi kebangkrutan suatu perusahaan, namun kebanyakan perusahaan kurang maksimal dalam mengelola aspek tersebut. Dalam memprediksi kebangkrutan perusahaan terdapat beberapa fitur, namun fitur yang tidak relevan dapat menyebabkan penurunan kinerja algoritma *machine learning*.

Memprediksi kebangkrutan perusahaan tidak lepas dari pengelolaan variabel yang telah dipilih melalui seleksi fitur, peningkatan akurasi prediksi kebangkrutan perusahaan dengan model algoritma C4.5 berbasis *forward selection* menjadi tujuan dari penelitian ini sehingga diharapkan dapat memberikan pengetahuan positif bagi pengembangan metode yang berkaitan dengan prediksi kebangkrutan perusahaan.

2. Metode

2.1 Pengumpulan Data

Model penelitian yang dilakukan adalah penelitian eksperimen, Data yang digunakan pada penelitian ini merupakan dataset global, dataset didonorkan oleh Dr. V. Prasanna Venkatesan, tanggal dipublikasikan pada Februari 2014. Dataset digunakan terakhir oleh Myoung-Jong Kim, Ingoo Han. 2003. "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms". *Expert System With Application*, 25, 637 – 646. [8].

Data tersebut merupakan hasil ekstraksi keputusan ahli, ahli menggunakan pengetahuan subyektif mereka untuk mendapatkan informasi kualitatif dengan langkah analisis sebagai berikut [8] :

1. Investigasi pendahuluan
2. Analisis laporan keuangan
3. Analisis tingkat industri
4. Analisis tingkat perusahaan
5. Evaluasi strategi pembiayaan
6. Estimasi resiko tetap.

2.2 Desain Eksperimen

Hasil pengumpulan data yang telah dimasukkan kedalam Ms. Excel selanjutnya diuji menggunakan algoritma C4.5 dan algoritma C4.5 berbasis *forward selection*, proses uji dataset dilakukan dengan memanfaatkan *tools Rapid Miner*.

Proses pelatihan dalam algoritma C4.5 :

1. Tentukan akar dari pohon dengan menggunakan *gain ratio*. Atribut dengan nilai *gain ratio* tertinggi dipilih sebagai atribut test untuk simpul. Langkah menghitung *gain ratio* :
 - a. Hitung nilai total *entropy* data set dan masing – masing atribut
 - b. Hitung nilai gain masing – masing atribut, dengan metode *information gain*.
 - c. Hitung *split info* masing – masing atribut.
 - d. Setelah mendapatkan nilai diatas hitung *gain ratio* semua atribut.
2. Menentukan Leaf Node berikutnya :
 - a. Label yang mengandung nilai Positif (+) dan Negatif (-) merupakan leaf node berikutnya
 - b. Node yang terpilih merupakan node yang memiliki nilai *gain ratio* tertinggi
 - c. Apabila node memiliki nilai *gain ratio* yang sama maka akan ditentukan oleh pakar (expert)
3. Lakukan langkah yang sama untuk leaf node berikutnya
4. Setelah pohon keputusan terbentuk langkah selanjutnya adalah membuat rule berdasarkan pohon keputusan yang telah ada
5. Aturan – aturan yang diperoleh dari pohon keputusan tersebut untuk membentuk model prediksi kebangkrutan perusahaan
6. Apabila terdapat hasil data yang bertentangan pada proses ini maka diperlukan seorang pakar untuk pengambilan keputusan.

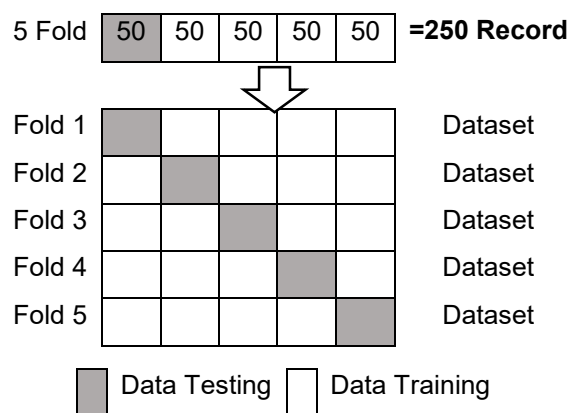
Proses eksperimen dilanjutkan dengan menggunakan algoritma *forward selection* yang berguna untuk mengurangi dimensi data, menghilangkan data yang tidak relevan, serta meningkatkan hasil akurasi. Proses pelatihan Algoritma C4.5 Berbasis *Forward Selection* :

1. Mulai dengan tidak ada *predictor* variabel (model hanya berisi konstanta).

2. Untuk semua *predictor* variabel tidak dalam model, pilih satu variabel dengan nilai p-value terkecil dan kurang dari taraf nyata α .
3. Ulangi langkah b, hingga tidak terdapat *predictor* variabel yang dapat ditambahkan ke dalam model.

Setelah variabel seleksi, langkah selanjutnya pengujian data set hasil seleksi menggunakan algoritma C4.5.

Data set akan dikelompokkan secara acak terdiri dari data uji 250 record, 150 record dan 100 record untuk menentukan atribut yang signifikan. Eksperimen yang dilakukan menggunakan *tools rapid miner* dengan teknik *cross validation*, data akan dibagi dalam n buah partisi dengan ukuran yang sama $D_1, D_2, D_3, \dots, D_n$ selanjutnya proses *testing* dan *training* dilakukan sebanyak n kali, dalam iterasi ke- i partisi D_i akan menjadi data *testing* dan sisanya akan menjadi data *training*. Dalam berapa kali *cross validation* menset pelatihan dibagi menjadi beberapa bagian yang sama besar yaitu dua subproses, subproses pelatihan dan subproses pengujian. Subproses pelatihan digunakan untuk pelatihan model. Model dilatih kemudian diterapkan dalam subproses pengujian. Pembagian dataset dalam proses 5-fold *cross validation* seperti terlihat pada gambar 1.



Gambar 1. Iterasi data dengan 5-fold *cross validation*

Hasil eksperimen yang dilakukan akan di evaluasi, model yang terbentuk diuji dengan menggunakan *confusion matrix* untuk mengetahui tingkat akurasi.

2.3 Proses Evaluasi

Hasil eksperimen yang dilakukan akan di evaluasi, diperlukan cara yang sistematis untuk mengevaluasi hasil. Evaluasi klasifikasi didasarkan pada pengujian pada obyek benar dan salah. Urutan pengujian ditabulasikan dalam *confusion matrix* dimana kelas yang diprediksi ditampilkan dibagian atas matriks dan kelas yang diamati disisi kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi.

Model yang terbentuk diuji dengan menggunakan *confusion matrix* untuk mengetahui tingkat akurasi. *Confusion matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang positif (True Positif), prediksi positif yang negatif (False Negatif), prediksi negatif yang positif (False Positif), prediksi negatif yang negative (False Negatif). Akurasi akan dihitung dari seluruh prediksi yang benar (baik positif dan negatif) dibandingkan dengan seluruh data testing, semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.

3. Hasil dan Pembahasan

3.1 Hasil Algoritma C4.5

Dalam penelitian ini data set yang digunakan terdiri dari 250 record dengan 6 atribut *regular* dan 1 atribut *special* sebagai target, berikut contoh data set yang digunakan dalam penelitian ini.

Tabel 1. Dataset Perusahaan

IR	MR	FF	Credibility	Compettiveness	Operation Risk	Bangkrut (Class)
Positif	Positif	Rata-Rata	Rata-Rata	Rata-Rata	Positif	Tidak

IR	MR	FF	Credibility	Compettiveness	Operation Risk	Bangkrut (Class)
Negatif	Negatif	Rata-Rata	Rata-Rata	Rata-Rata	Negatif	Tidak
Rata-Rata	Rata-Rata	Rata-Rata	Rata-Rata	Rata-Rata	Rata-Rata	Tidak
Positif	Positif	Positif	Positif	Positif	Positif	Tidak
Negatif	Negatif	Positif	Positif	Positif	Negatif	Tidak
Rata-Rata	Rata-Rata	Positif	Positif	Positif	Rata-Rata	Tidak
Positif	Positif	Rata-Rata	Positif	Positif	Positif	Tidak
Positif	Positif	Positif	Rata-Rata	Rata-Rata	Positif	Tidak
Positif	Positif	Rata-Rata	Positif	Rata-Rata	Positif	Tidak
Positif	Positif	Rata-Rata	Rata-Rata	Positif	Positif	Tidak
Positif	Positif	Positif	Positif	Rata-Rata	Positif	Tidak
Positif	Positif	Positif	Rata-Rata	Positif	Positif	Tidak
Negatif	Negatif	Positif	Rata-Rata	Rata-Rata	Negatif	Tidak
Negatif	Negatif	Negatif	Rata-Rata	Negatif	Positif	Ya

Tahap dalam membuat pohon keputusan dengan algoritma C4.5 :

1. Menentukan akar dari pohon keputusan, akar diambil dari atribut yang terpilih dengan cara menghitung nilai *gain ratio* dari masing – masing atribut, nilai *gain ratio* yang paling tinggi akan menjadi akar pertama.
2. Menghitung nilai *entropy* total dengan menggunakan persamaan berikut :

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

$$Entropy \text{ Total} = -\frac{143}{250} \log_2 \frac{143}{250} + -\frac{107}{250} \log_2 \frac{107}{250} = 0.98$$

Dimana terdapat 143 data perusahaan tidak bangkrut dan 107 data perusahaan bangkrut.

3. Lakukan perhitungan *entropy* untuk masing – masing atribut, untuk perhitungan atribut resiko industri (IR) tetap menggunakan persamaan (1) dengan penjelasan sebagai berikut :

Tabel 2. Perhitungan Kasus Atribut IR

IR	Bangkrut	JUMLAH
Positif	YA (+)	26
Positif	TIDAK (-)	54
Negatif	YA (+)	53
Negatif	TIDAK (-)	36
Rata – Rata	YA (+)	28
Rata – Rata	TIDAK (-)	53

$$\text{Positif} = -\frac{54}{80} \log_2 \frac{54}{80} + -\frac{26}{80} \log_2 \frac{26}{80} = 0.91$$

$$\text{Negatif} = -\frac{53}{89} \log_2 \frac{53}{89} + -\frac{36}{89} \log_2 \frac{36}{89} = 0.97$$

$$\text{Rata - Rata} = -\frac{28}{81} \log_2 \frac{28}{81} + -\frac{53}{81} \log_2 \frac{53}{81} = 0.93$$

Setelah mendapatkan hasil *entropy* dari masing – masing nilai atribut IR langkah selanjutnya adalah menghitung nilai gain atribut IR menggunakan persamaan berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

$$Gain(Total, IR) = 0.98 - \left(\left(\frac{80}{250} * 0.91 \right) + \left(\frac{89}{250} * 0.97 \right) + \left(\frac{81}{250} * 0.93 \right) \right) = 0.042$$

4. Perhitungan nilai *split info* menggunakan persamaan sebagai berikut :

$$SplitInfo(S, A) = - \sum_{i=1}^c \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

$$Split\ info\ (Total, IR) = - \frac{80}{250} \log_2 \frac{80}{250} + - \frac{89}{250} \log_2 \frac{89}{250} + - \frac{81}{250} \log_2 \frac{81}{250} = 1.58$$

5. Menghitung *gain ratio* untuk menentukan atribut sebagai akar dari pohon keputusan menggunakan persamaan 2.4

$$Gain\ Ratio\ (S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (4)$$

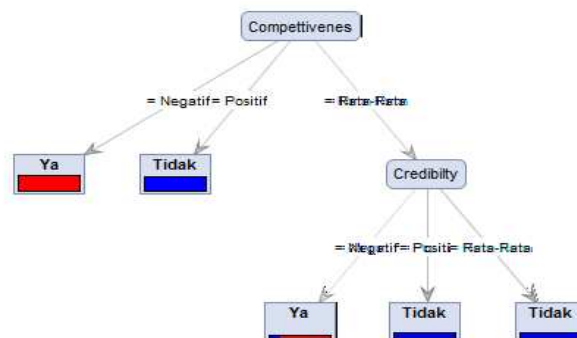
$$Gain\ Ratio\ (Total, IR) = \frac{0.042}{1.58} = 0.0266$$

Untuk atribut yang lain juga dilakukan perhitungan yang sama, nilai *gain ratio* yang paling tinggi akan menjadi node akar dari pohon. Dari hasil yang didapatkan nilai *gain ratio* tertinggi dimiliki oleh atribut *compettivines* yaitu 0.5820, sehingga ditetapkan *compettivines* menjadi akar dari pohon keputusan sesuai dengan tabel 3 dibawah ini :

Tabel 3. Perhitungan Node 1

ATRIBUT	GAIN RATIO
IR	0.0266
MR	0.0676
FF	0.3830
CREDIBILITY	0.3358
COMPETTIVINES	0.5820
OPERATION RISK	0.03423

Atribut *compettivines* memiliki tiga buah nilai positif, negatif, dan rata – rata, dari nilai tersebut positif sudah mengklasifikasikan kasus menjadi tidak, dan nilai negatif mengklasifikasikan kasus menjadi ya, sehingga tidak perlu dilakukan partisi node lagi. Untuk nilai rata – rata perlu dilakukan partisi node lagi. Untuk mendapatkan node selanjutnya lakukan perhitungan yang sama untuk 5 atribut lainnya sehingga kelima nilai atribut mengklasifikasikan kasus menjadi 1 keputusan dan membentuk pohon keputusan seperti gambar berikut:



Gambar 2. Pohon Keputusan Algoritma C4.5 Dengan 250 Record

Pohon keputusan dari gambar 4.1 dapat diekstraksi sejumlah aturan sebagai berikut :

- R1 : IF Compettiveness = Positif Then Bangkrut = Tidak
- R2 : IF Compettiveness = Negatif Bangkrut = Ya
- R3 : IF Compettiveness = Rata - Rata dan Credibility = Positif Then Bangkrut = Tidak
- R4 : IF Compettiveness = Rata - Rata dan Credibility = Negatif Then Bangkrut = Ya
- R5 : IF Compettiveness = Rata - Rata dan Credibility = Rata - Rata Then Bangkrut = Tidak

Model yang terbentuk diuji dengan menggunakan *confusion matrix* untuk mengetahui tingkat akurasi. *Confusion matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang positif (True Positif), prediksi positif yang negatif (False Negatif), prediksi negatif yang positif (False Positif), prediksi negatif yang negative (False Negatif). Untuk menghitung nilai *accuracy*, *precision*, dan *recall* dengan rumus perhitungan sebagai berikut :

$$Accuracy = \left(\frac{a+d}{a+b+c+d} \right) = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

Keterangan :

- TP = Prediksi positif yang positif
- FN = Prediksi positif yang negatif
- FP = Prediksi negatif yang positif
- TN = Prediksi negatif yang negatif

Eksperimen ini menggunakan data sebanyak 250 record, berdasarkan hasil *confusion matrix* terlihat bahwa 107 record pada class a (prediksi positif yang positif), 1 record pada class b (prediksi positif yang negative), 0 record pada class c (prediksi negatif yang positif) dan 142 record pada class d (prediksi negatif yang negatif).

$$Accuracy = \left(\frac{107+142}{107+1+0+142} \right) = 0.996$$

$$Precision = \frac{107}{107+0} = 1$$

$$Recall = \frac{107}{107+1} = 0.99$$

Untuk mendapatkan presentase masing – masing perhitungan diatas hasil tersebut akan dikalikan dengan 100 sehingga didapatkan nilai *recall* sebesar 99%, *precision* sebesar 100% dan *accuracy* sebesar 99.60%. Dalam penelitian ini eksperimen menggunakan algoritma C4.5 dilakukan dengan variasi jumlah data yang digunakan, berikut hasil akurasi dalam eksperimen selanjutnya :

Tabel 4. Hasil Pengujian Algoritma C4.5

Jumlah Record	Accuracy
250	99.60%
150	98.67%
100	99.02%

3.2 Hasil Algoritma C4.5 Berbasis *Forward Selection*

Hasil pelatihan algoritma C4.5 berbasis *forward selection* menggunakan data set yang sama dengan 250 record memberikan 2 atribut terbaik yang bernilai 1, dan 4 atribut bernilai 0 akan dibuang. Atribut hasil seleksi fitur adalah *Credibility* dan *Competiveness* dengan tingkat akurasi sebesar 99.61%.

Tabel 5. Hasil Seleksi Fitur Dengan 250 Record

Atribut	Weight
IR	0
MR	0
FF	0
Credibility	1
Compettiveness	1
Operation Risk	0

Penelitian algoritma C4.5 berbasis *forward selection* menggunakan data set yang sama dengan 150 record dengan 6 atribut *regular* dan 1 atribut *special* sebagai target, terdiri dari 55 perusahaan bangkrut dan 95 perusahaan tidak bangkrut memberikan 2 hasil terbaik dalam pengujian, pada validasi 7 dan 9 menghasilkan nilai akurasi, *recall* dan *precision* yang sama namun atribut hasil seleksi fitur yang dihasilkan berbeda pada salah satu atribut. Pada validasi 7 atribut hasil seleksi fitur adalah *Compettiveness*, MR dan *Credibility* sementara pada validasi 9 atribut hasil seleksi fitur adalah *Compettiveness*, IR dan *Credibility*.

Penelitian algoritma C4.5 berbasis *forward selection* menggunakan data set yang sama dengan 100 record dengan 6 atribut *regular* dan 1 atribut *special* sebagai target, terdiri dari 45 perusahaan bangkrut dan 55 perusahaan tidak bangkrut. Hasil evaluasi dan validasi memberikan 2 atribut terbaik yang bernilai 1, dan 4 atribut bernilai 0, atribut hasil seleksi fitur adalah *Compettiveness* dan IR. Dengan variasi jumlah data *training* didapatkan hasil yang signifikan, berikut hasil pengujian algoritma C4.5 berbasis *forward selection*.

Tabel 6. Hasil Pengujian Algoritma C4.5 Berbasis *Forward Selection*

Jumlah Record	Accuracy	Seleksi Fitur
250	99.61%	Credibility, Compettiveness
150	99.35%	Credibility, MR dan Compettiveness
100	99.07%	Compettiveness dan IR

3.3 Pembahasan

Dari hasil penelitian yang telah dilakukan dengan penerapan algoritma C4.5 dan algoritma C4.5 berbasis *forward selection* dalam prediksi kebangkrutan perusahaan menunjukkan peningkatan akurasi yang baik. Dalam penelitian dengan 250 record algoritma C4.5 mendapatkan hasil akurasi sebesar 99.60% dan algoritma C4.5 berbasis *forward selection* dengan akurasi sebesar 99.61%.

Sementara penelitian menggunakan 150 record algoritma C4.5 mendapatkan akurasi sebesar 98.67% dan algoritma C4.5 berbasis *forward selection* menghasilkan akurasi sebesar 99.35%, untuk penelitian menggunakan 100 record hasil dari algoritma C4.5 sebesar 99.02% dan algoritma C4.5 berbasis *forward selection* menghasilkan tingkat akurasi lebih baik dengan akurasi sebesar 99.07%. Seleksi fitur *forward selection* menghasilkan beberapa fitur signifikan, dalam penelitian menggunakan 250 record atribut *compettiveness* dan *credibility* menjadi fitur yang signifikan dari 6 atribut yang ada, penelitian menggunakan 150 record terdapat 2 hasil seleksi atribut yang signifikan, dimana pada validasi 7 atribut signifikan *Compettiveness*, MR dan *Credibility* dan pada validasi 9 atribut signifikan *Compettiveness*, IR dan *Credibility*, penelitian menggunakan 100 record menghasilkan atribut *Compettiveness* dan IR dari 6 atribut yang ada.

Dengan variasi jumlah record dalam penelitian ini atribut yang paling signifikan terdapat pada atribut *Compettiveness* dan *Credibility*, hal tersebut dapat dilihat dari penelitian menggunakan 250 record dan 150 record, atribut IR signifikan dalam penelitian menggunakan 150 record dan 100 record, sementara atribut MR hanya signifikan pada penelitian menggunakan data 150 record.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian dan pembahasan, maka dapat disimpulkan bahwa :

4.1 Kesimpulan

Dari penelitian yang telah dilakukan dapat disimpulkan beberapa hal yaitu :

1. Dengan variasi jumlah record dalam penelitian ini atribut yang paling signifikan terdapat pada atribut *Compettiveness* dan *Credibility*, hal tersebut dapat dilihat dari penelitian menggunakan 250 record dan 150 record, atribut IR signifikan dalam penelitian menggunakan 150 record dan 100 record, sementara atribut MR hanya signifikan pada penelitian menggunakan data 150 record. Dengan demikian pengelolaan variabel hasil seleksi fitur dapat dioptimalkan agar dapat menghindari kebangkrutan perusahaan.
2. Hasil penelitian dengan variasi jumlah record yang telah dilakukan dengan algoritma C4.5 dan algoritma C4.5 berbasis *forward selection* dalam prediksi kebangkrutan perusahaan menunjukkan teknik seleksi fitur *forward selection* dapat menghapus fitur yang tidak relevan sehingga akurasi algoritma C4.5 mengalami peningkatan yang baik. Dengan peningkatan akurasi yang didapatkan model algoritma C4.5 berbasis *forward selection* dapat menjadi kontribusi untuk perusahaan dalam memprediksi kebangkrutan.

4.2 Saran

Berdasarkan hasil penelitian dan pembahasan, maka beberapa saran yang dapat diberikan adalah sebagai berikut :

1. Penggunaan teknik lain dalam memprediksi kebangkrutan perusahaan sangat dibutuhkan, sehingga dapat menghasilkan pola prediksi kebangkrutan perusahaan yang lebih baik.
2. Hasil prediksi kebangkrutan perusahaan dapat menjadi kontribusi untuk perusahaan namun terdapat beberapa hal yang perlu diperhatikan terkait dengan data set bertipe *ordinal* (positif, negatif dan rata – rata) karena membutuhkan analisa ahli dalam penentuan nilai tersebut.

Daftar Pustaka

- [1] Li H Andina D & Sun J. 2012. "Multiple proportion case-basing driven CBRE and its application in the evaluation of possible failure of firms". *International Journal of Systems Science*, 44, 1409–1425.
- [2] Olson D. L, Delen D & Meng Y. 2012. "Comparative analysis of data mining methods for bankruptcy prediction". *Decision Support Systems*, 52, 464–473.
- [3] Gang Wang, Jinxing Hao, Jian Ma & Hongbing Jiang. 2011. "A comparative assessment of ensemble learning for credit scoring". *Expert Systems with Applications*, 38, 223–230.
- [4] Duéñez-Guzmán, E. A, & Vose, M. D. 2013. "No free lunch and benchmarks". *Evolutionary Computation*, 21, 293–312.
- [5] Myoung-Jong Kim, Dae-Ki Kang. 2012. "Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction". *Expert System With Application*, 39, 9308 – 9314.
- [6] Matthias Reif, Faisal Shahait, 2014. "Efficient Feature Size Reduction Via Predictive Forward Selection". *Pattern Recognition*, 47, 1664–1673.
- [7] L. Ladha, T. Deepa. 2011. "Feature Selection Methods And Algorithms", *International Journal on Computer Science and Engineering (IJCSE)*, ISSN : 0975-3397 Vol. 3 No.
- [8] Myoung-Jong Kim, Ingoo Han. 2003. "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms". *Expert System With Application*, 25, 637 – 646.