

MultiClass Decision Forest Machine Learning Artificial Intelligence

Bambang Siswoyo¹

* Departemen Information system, STKOM AI-Ma'soem Jatinangor JABAR
bambangfl@gmail.com

Article Info

Article history:

Received 2019-03-06

Revised 2019-11-22

Accepted 2020-01-21

Keyword:

*Kecerdasan buatan,
MultiClass Decision Forest,
Machine Learning,
Pemodelan.*

ABSTRACT

Kecerdasan buatan merupakan solusi yang terbaik dalam menangani berbagai bidang. Algoritma Multi Class Decision Forest Machine Learning yang merupakan bagian dari kecerdasan buatan adalah bidang yang menarik untuk diterapkan dalam bidang perbankan. Penelitian ini dikembangkan untuk membangun model yang dapat memprediksi dan mengevaluasi kebangkrutan industri perbankan. Variabel prediktor adalah rasio keuangan yang didapatkan dari publikasi situs <http://www.idx.co.id>. Modeling machine learning dengan enam variabel, dimana lima variabel sebagai input dan satu variabel sebagai target. Secara keseluruhan, Multi Class Decision Forest Machine Learning mampu melatih data hubungan input-output dan perilaku pemodelan dengan baik, nilai accuracy 92%, nilai precision 92% dan nilai under area curve 90%.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kebangkrutan biasanya diartikan sebagai kegagalan perusahaan dalam menjalankan operasi perusahaan untuk menghasilkan laba. Dengan kata lain, ini terkait dengan kinerja organisasi bisnis seperti perbankan untuk mempertahankan laba. Kebangkrutan adalah urusan bisnis yang menarik untuk dipelajari. Perbankan adalah denyut nadi perekonomian di seluruh negara. Perbankan di Indonesia memiliki peran yang sangat penting, salah satunya adalah menjaga stabilitas moneter yang disebabkan oleh kebijakannya terhadap tabungan publik serta lalu lintas pembayaran. Ada banyak metode yang telah diterapkan saat ini untuk mengevaluasi dan menganalisis kinerja perbankan menggunakan analisis laporan keuangan. Laporan keuangan adalah ringkasan perusahaan yang dapat digunakan sebagai pengetahuan prediktif [1][2] [3].

Metode decision tree membangun model keputusan yang dibuat berdasarkan nilai actual atribut dalam data. Penggunaan pohon (tree) yang semakin banyak maka, mempengaruhi akurasi yang didapatkan, yakni akan menjadi lebih baik. Penentuan klasifikasi dengan random forest diambil berdasarkan hasil voting dari tree tertentu [4].

Machine learning adalah bagian dari kecerdasan buatan, yang mempelajari pengenalan pola untuk mendapatkan solusi yang optimal. Machine learning memungkinkan komputer menemukan solusi data yang optimal secara otomatis [5]. Pembelajaran mesin telah terbukti menyelesaikan banyak

tugas. Pembelajaran mesin untuk memprediksi sebelumnya telah digunakan dalam konteks memprediksi kebangkrutan dengan SVM [6] dan Robust Logistic Regression [7], model ini memiliki keterbatasan. Dimana nilai teknik ini sangat bergantung pada kemampuan peneliti untuk memasukkan variabel independen yang benar.

Dengan kata lain, jika peneliti gagal mengidentifikasi semua variabel independen yang relevan, regresi logistik memiliki nilai prediksi yang rendah [8]. Akurasi keseluruhan adalah 75,69% di set pelatihan (training) dan 69,44% di set pengujian (testing). Penggunaan algoritma backpropagation sebagai model prediksi kebangkrutan bank-bank komersial di Indonesia [9] [10], dengan model prediksi kepailitan bank umum di Indonesia menggunakan Backpropagation.

Dari sini menjadi jelas perlunya model alternatif tetapi juga metode yang dapat diandalkan untuk memprediksi dan menganalisis kinerja bisnis perbankan diperlukan. Studi ini mengusulkan penggunaan pendekatan Algoritma Boosted Decision Tree untuk prediksi kinerja perbankan.

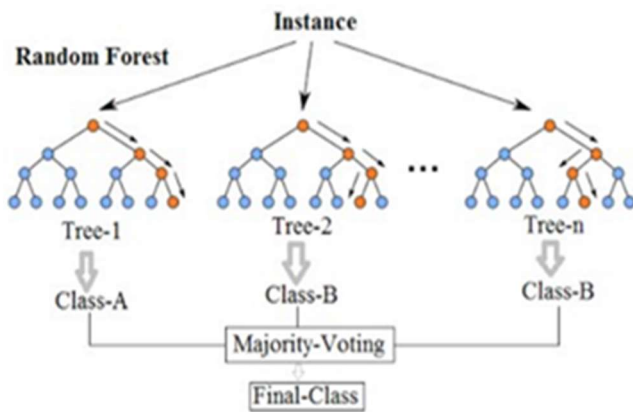
II. LANDASAN TEORI

A. Multi Class Decision Forest

Algoritma decision forest adalah metode pembelajaran ensemble untuk klasifikasi. Algoritma ini bekerja dengan membangun beberapa pohon keputusan dan kemudian memilih kelas output yang paling populer. Voting adalah suatu bentuk agregasi, dimana setiap pohon dalam hutan

keputusan klasifikasi menghasilkan histogram frekuensi label yang tidak dinormalisasi. Proses agregasi menjumlahkan histogram ini dan menormalkan hasilnya untuk mendapatkan "probabilitas" untuk setiap label.

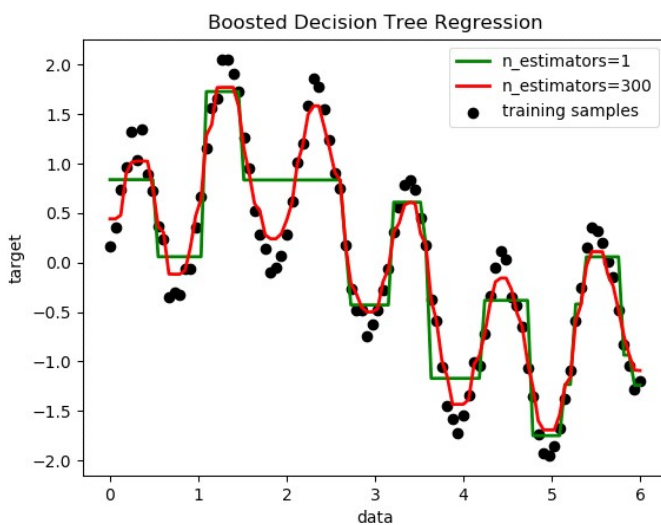
Pohon-pohon yang memiliki kepercayaan prediksi tinggi memiliki bobot lebih besar dalam keputusan akhir ansambel. Pohon keputusan secara umum adalah model non-parametrik, yang berarti mereka mendukung data dengan distribusi yang bervariasi. Di setiap pohon, serangkaian tes sederhana dijalankan untuk setiap kelas, meningkatkan tingkat struktur pohon hingga simpul daun (keputusan) tercapai [11]. Ilustrasi digambarkan pada gambar 1 sebagai berikut.



Gambar 1. Multiclass Decision Forest

B. Boosted Decision Tree Regression.

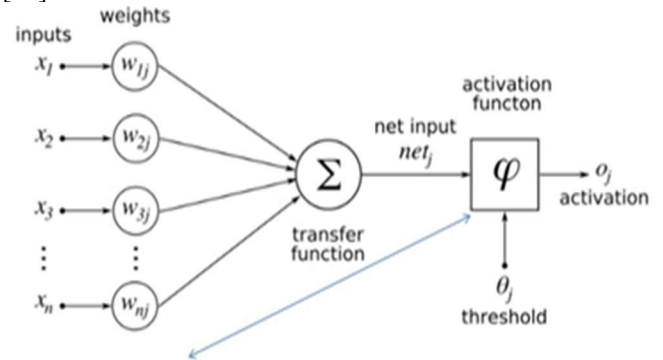
Boosted Decision Tree Regression menggunakan algoritma AdaBoost. Selama beberapa tahun terakhir, teknik ini telah muncul sebagai salah satu metode yang paling kuat untuk menambang data prediksi. Beberapa implementasi dari algoritma yang kuat ini memungkinkan mereka untuk digunakan untuk regresi serta masalah klasifikasi [12] [13]. Ilustrasi digambarkan pada gambar 2 sebagai berikut.



Gambar 2. Boosted Decision Tree Regression

C. Neural Network Regression

Secara umum, jaringan saraf digunakan untuk tujuan pengelompokan melalui pembelajaran terawasi, atau regresi [14].



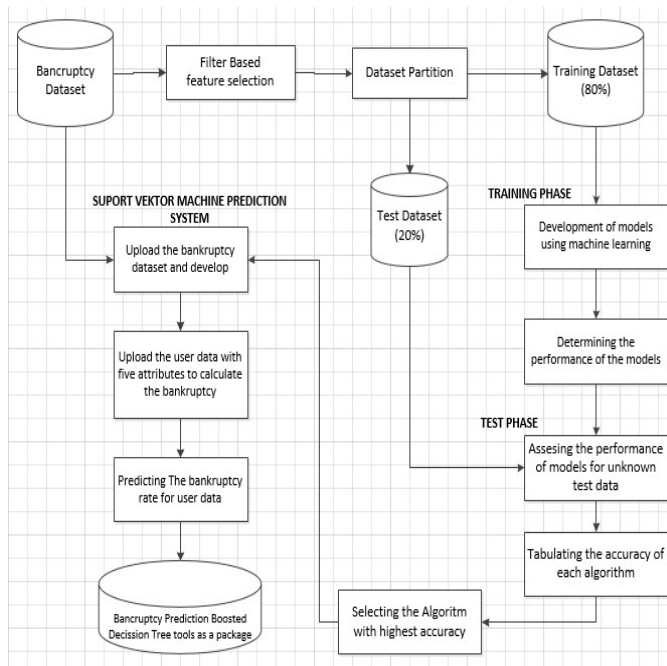
Gambar 3. Neural Network Regression

Pada gambar 3 diagram di atas, x singkatan input, fitur diteruskan ke depan dari lapisan jaringan sebelumnya. Banyaknya x dimasukkan ke dalam setiap node dari lapisan terakhir yang tersembunyi, dan setiap x akan dikalikan dengan bobot yang sesuai, (w). Jumlah produk ditambahkan ke bias dan dimasukkan ke dalam fungsi aktivasi. Dalam hal ini fungsi aktivasi adalah unit linier yang ditingkatkan (ReLU), yang umum digunakan dan sangat berguna seperti fungsi aktivasi sigmoid.

Jaringan saraf yang mengalami regresi akan memiliki satu simpul keluaran. Untuk melakukan backpropagasi [15] [16] dan membuat jaringan pembelajaran, cukup dengan membandingkan \hat{y} dengan nilai ground-truth y dan menyesuaikan bobot dan bias jaringan hingga kesalahan diminimalkan. Dengan cara ini, maka dapat menggunakan jaringan saraf untuk mendapatkan fungsi yang terkait dengan sejumlah variabel independen yang berubah x menjadi variabel dependen y yang akan diprediksi.

III. METODE PENELITIAN

Data time series dari lima variabel rasio keuangan bank yang diterbitkan di Bank Indonesia dari 2011 hingga 2015 telah dikumpulkan dan dianalisis. Pada tahap preprocessing, data deret waktu kemudian dikonversi menjadi matriks ukuran [62x6] (62 baris dan 6 kolom). Dari matriks [123x5] telah dibagi menjadi dua bagian yaitu untuk data pelatihan dan data pengujian, sebagai input ke dalam model. Data input terdiri dari lima variable independent, variable modal kerja terhadap total aset, rasio laba ditahan terhadap total aset, penghasilan sebelum bunga dan pajak terhadap total aset, nilai pasar ekuitas dan nilai buku terhadap total kewajiban, penjualan terhadap total aset. Sedangkan outputnya adalah klasifikasi bankrupt, gray area dan non bankrupt.



Gambar 4. Metode Penelitian

Metode penelitian yang digunakan mulai dari menyiapkan dataset kebangkrutan perbankan yang mempublikasikan laporan keuangannya di BEI, berupa rasio keuangan sejumlah 6 rasio. Memilih algoritma Boosted Decision Tree, membagi data menjadi dua yaitu data training 80% dan data testing 20%, setelah melakukan data training kemudian data testing, hasil dari algoritma tersebut di validasi dengan koefisien regresi dan root means square error (RMSE).

Adapun tahapan implementasi adalah sebagai berikut.

- 1) *Bankruptcy dataset*, menyiapkan dataset kebangkrutan untuk input sistem klasifikasi.
- 2) *Filter based feature selection*, sebuah fungsi untuk memilih berdasarkan feature yang terbaik.
- 3) *Dataset partion*, melakukan pembagian data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Training dataset, data pelatihan sebanyak 80%, sedangkan data pengujian sebanyak 20%.
Data pelatihan terdiri dari pengembangan model menggunakan machine learning dan penentuan kinerja model (output data pelatihan).
Data pengujian terdiri dari penilaian kinerja model untuk data uji yang tidak diketahui dan tabulasi keakuratan masing-masing algoritma (output data pengujian).
- 4) *Pemilihan Algoritma*
Pemilihan algoritma dengan akurasi tertinggi yang dipilih.
- 5) *Mengunggah dan mengembangkan data set kebangkrutan*.
- 6) *Mengunggah data pengguna dengan lima atribut untuk menghitung kebangkrutan*.

- 7) *Prediksi tingkat kebangkrutan untuk data pengguna*.
- 8) *Prediksi kebangkrutan menggunakan Decision Tree Regression Machine sebagai paket*. Algoritma yang digunakan dalam klasifikasi kebangkrutan bank adalah metode machine learning (Neural network, Multi class decision forest).

Dalam penelitian ini diusulkan Multi Class Decision Forest, dengan menggunakan aplikasi Microsoft Azure. Variabel dan sumber data yang digunakan dalam penelitian ini dapat dilihat pada table 1 sebagai berikut.

TABEL I
VARIABEL DAN SUMBER DATA

No	Rasio Keuangan	Variable	Sumber
1	<i>Working capital to total asset</i>	X1	http://www.idx.co.id
2	<i>Retained earning to total asset</i>	X2	http://www.idx.co.id
3	<i>Earning before interest and taxed to total asset</i>	X3	http://www.idx.co.id
4	<i>Market value of Equity to book value of total liabilities</i>	X4	http://www.idx.co.id
5	<i>Sales to total asset</i>	X5	http://www.idx.co.id

IV. HASIL DAN PEMBAHASAN

A. Dataset Rasio Keuangan

Rasio keuangan, adalah yang menghubungkan dua data keuangan dengan jalan membagi satu data dengan data yang lain. Indeks yang menghubungkan dua angka akuntansi dan diperoleh dengan membagi satu angka dengan angka yang lainnya. Perhitungan rasio digunakan karena cara ini akan diperoleh perbandingan yang lebih berguna daripada melihat angka saja. Analisa rasio keuangan melibatkan dua jenis perbandingan.

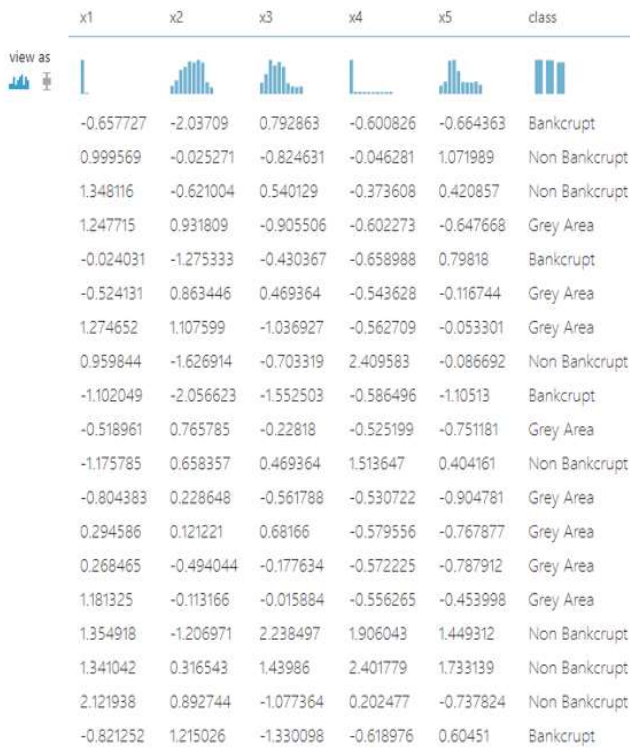
Pertama, analisis dapat membandingkan rasio saat ini dengan rasio masa lalu dan akan datang dalam perusahaan yang sama. Rasio lancar untuk tahun sekarang dapat dibandingkan dengan rasio lancar tahun sebelumnya. Jika rasio keuangan diurutkan dalam beberapa periode tahun, analisis dapat mempelajari komposisi perubahan dan menentukan apakah terdapat perbaikan atau penurunan dalam kondisi keuangan dan kinerja perusahaan.

Metode perbandingan kedua melibatkan perbandingan rasio satu perusahaan dengan perusahaan-perusahaan sejenis atau dengan rata-rata industri pada titik waktu yang sama. Perbandingan ini memberikan pandangan mendalam tentang kondisi keuangan dan kinerja relatif dari perusahaan [1] [2].

Hasil analisis menunjukkan bahwa semua lima atribut tersebut sangat terkait dengan variabel hasil. Dalam rangka untuk mengkonfirmasi hasil dari korelasi, metode seleksi fitur lain diterapkan untuk dataset. Informasi memperoleh metode filter peringkat diaplikasikan untuk menguji pentingnya fitur. Algoritma ditemukan hasil yang sama seperti yang dari

korelasi. Oleh karena itu semua lima atribut dari dataset dianggap perlu untuk penelitian. Nama dataset adalah indeks kebangkrutan industri perbankan berdasarkan 5 aspek rasio keuangan. Deskripsi dataset adalah dataset ini berisi indeks kebangkrutan industri perbankan berdasarkan 5 aspek rasio keuangan.

Dimana rasio finance (rasio keuangan) diambil dari laporan keuangan industri perbankan yang sudah di audit oleh auditor kemudian mempublikasinya di situs Bank Indonesia (<http://www.bi.go.id> atau <http://www.idx.co.id>). Plot korelasional untuk fitur ditunjukkan pada gambar 5.



Gambar 5. Data Rasio Keuangan

B. Bosted Decesion Tree Regression (BDTR)

Berdasarkan classifier BDFR sistem prediksi dibangun menggunakan software Microsoft A-Zure Machine Learning. Dataset rasio keuangan perbankan yang di publikasikan BI, awalnya dimuat ke dalam sistem prediktif sebagai file.csv, kemudian file.csv tersebut di simpan di penyimpanan lokal, selanjutnya system on line akan menempatkan dataset rasio keuangan dalam paper work kemudian system loading data tersebut untuk di proses dengan menggunakan algoritma yang diusulkan. Tampilan hasil prediksi dataset rasio keuangan industry perbankan ditunjukkan pada gambar 6.



Gambar 6. Hasil Bosted Decision Tree Regression

Gambar 6 menunjukkan hasil Boosted Decision Tree Regression, merupakan hasil algoritma klasifikasi, didapat hasil dengan true positive 20, falls negative 0, false positive 1, true negative 0, precision 95%, accuracy = 95% dan AUC =90%, ini menandakan bahwa akurasi mengacu pada seberapa dekat nilai yang diukur dengan nilai sebenarnya yang sebenarnya.

Perkiraan klasifikasi target mempunyai nilai perkiraan sangat dekat dengan nilai actual, presisi mengacu pada seberapa dekat nilai yang diukur satu sama lain. Jadi model BDTR sangat tepat karena mampu memperkirakan klaifikasi target actual hampir sama dengan klasifikasi hasil dari model saat meramalkan, memprediksi, atau memproyeksikan informasi yang kita inginkan dapat memberikan hasil yang akurat dan tepat.

Hal ini berarti bahwa secara konsisten mendapatkan hasil serupa yang semuanya sangat dekat dengan jawaban yang benar. Selain itu akan memberikan kepercayaan diri bahwa model itu tidak hanya benar tetapi secara konsisten benar, yang keduanya penting dari perspektif prediksi. AUC merupakan area di bawah kurva (Area under the Curve of), suatu kurva yang menggambarkan probabilitas dengan variabel sensitivitas dan kekhususan (specificity) dengan nilai batas antara 0 hingga 1. Area di bawah kurva memberikan gambaran tentang keseluruhan pengukuran atas kesesuaian

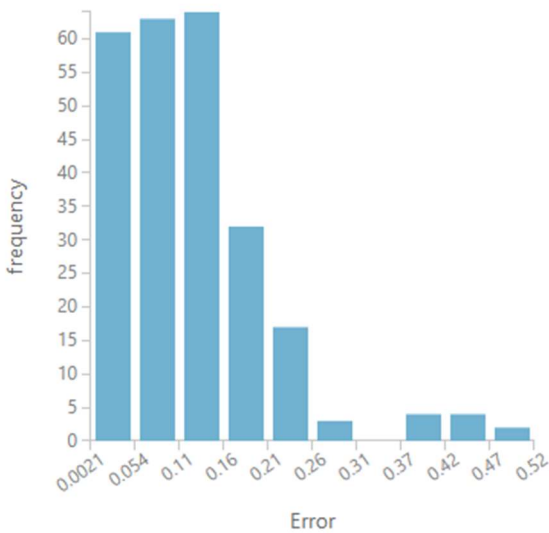
dari model yang digunakan. Model BDTR mempunyai kesesuaian 90% dengan system nyata.

C. Neural Network Regression

Metrics

Mean Absolute Error	0.119773
Root Mean Squared Error	0.148978
Relative Absolute Error	0.244618
Relative Squared Error	0.090657
Coefficient of Determination	0.909343

Error Histogram



Gambar 7. Hasil Neural Network

Gambar 7 merupakan hasil Neural Network Regression adalah sebagai berikut MAE 11%, RMSE 14%, Relative absolute error 24%, sedangkan Relative square error 24%. Coefesien of determination 90%.

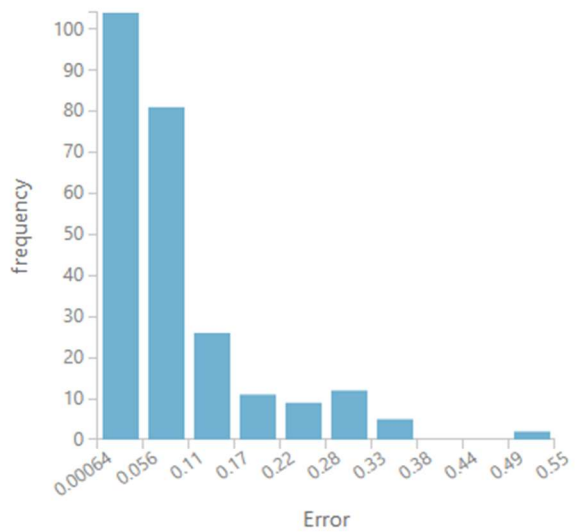
4.5 Multi Class Decision Forest (MCDF)

Berdasarkan classifier MCDF sistem prediksi dibangun di A-Zure. Kebangkrutan dataset awalnya dimuat ke dalam sistem prediktif sebagai file.csv. Tampilan hasil prediksi dataset keuangan kebangkrutan ditunjukkan pada gambar 8 sebagai berikut.

Metrics

Mean Absolute Error	0.09735
Root Mean Squared Error	0.133247
Relative Absolute Error	0.198824
Relative Squared Error	0.072523
Coefficient of Determination	0.927477

Error Histogram



Gambar 8. Hasil Class Decesion Forest Result

Gambar 8 menunjukkan bahwa hasil dari model MCDF adalah sebagai berikut: MAE 9%, RMSE 13%, Relative absolute error 19%, sedangkan Relative square error 7%. Coefesien of determination 92%.

Berdasarkan analisis komparatif, boosted decision menghasilkan nilai precision 95%, accuracy 95% dan AUC =90%, maka menandakan klasifikasi nilai perkiraan sangat tepat dengan nilai actual, sedangkan presisi 95% juga sangat dekat nilai yang diukur, nilai AUC BDTR mempunyai kesesuaian 90% dengan system nyata. Penelitian lebih lanjut menunjukkan bahwa boosted decision forest dapat digunakan untuk kasus berbeda lainnya tidak terbatas hanya dalam prediksi kebangkrutan.

TABLE II
REKAPITULASI HASIL ALGORITMA

No	Algoritma	MAE	RMSE	RAE	RSE	R
1	Bosted Decision Tree Regresion	0.03	0.10	0.04	0.01	0.98
2	Neural Network Regresion	0.11	0.14	0.24	0.09	0.90

3	Multi Class Decision Forest	0.09	0.13	0.19	0.07	0.92
---	-----------------------------	------	------	------	------	------

V. KESIMPULAN

Berdasarkan hasil diskusi berdasarkan hasil yang dibahas pada bab sebelumnya tentang pembentukan model prediksi kebangkrutan perbankan di Indonesia dengan menggunakan lima variabel input (modal kerja pada total aset, saldo laba pada total aset, penapatan bunga dan pajak atas total aset, nilai pasar Ekuitas terhadap total nilai buku liabilities, penjualan ke total aset) dalam menjelaskan output bangkrut (bankrupt) dengan metode pembelajaran mesin, maka beberapa kesimpulan dapat diambil dari analisis. Modeling Machine Learning dengan tiga algoritma dapat memodelkan perilaku input-output antara variable rasio keuangan dengan variable Bankrupt. Secara keseluruhan, Machine Learning (neural network, multiclass decision forest). Multi Class Decision Forest mampu melatih data dan memodelkan perilaku hubungan input-output dengan baik, ini dibuktikan dengan tingkat MAE 0,03 dan koefisien determinasi 0,92. Dari hasil prediksi ini, data pelatihan berhasil diuji dengan benar.

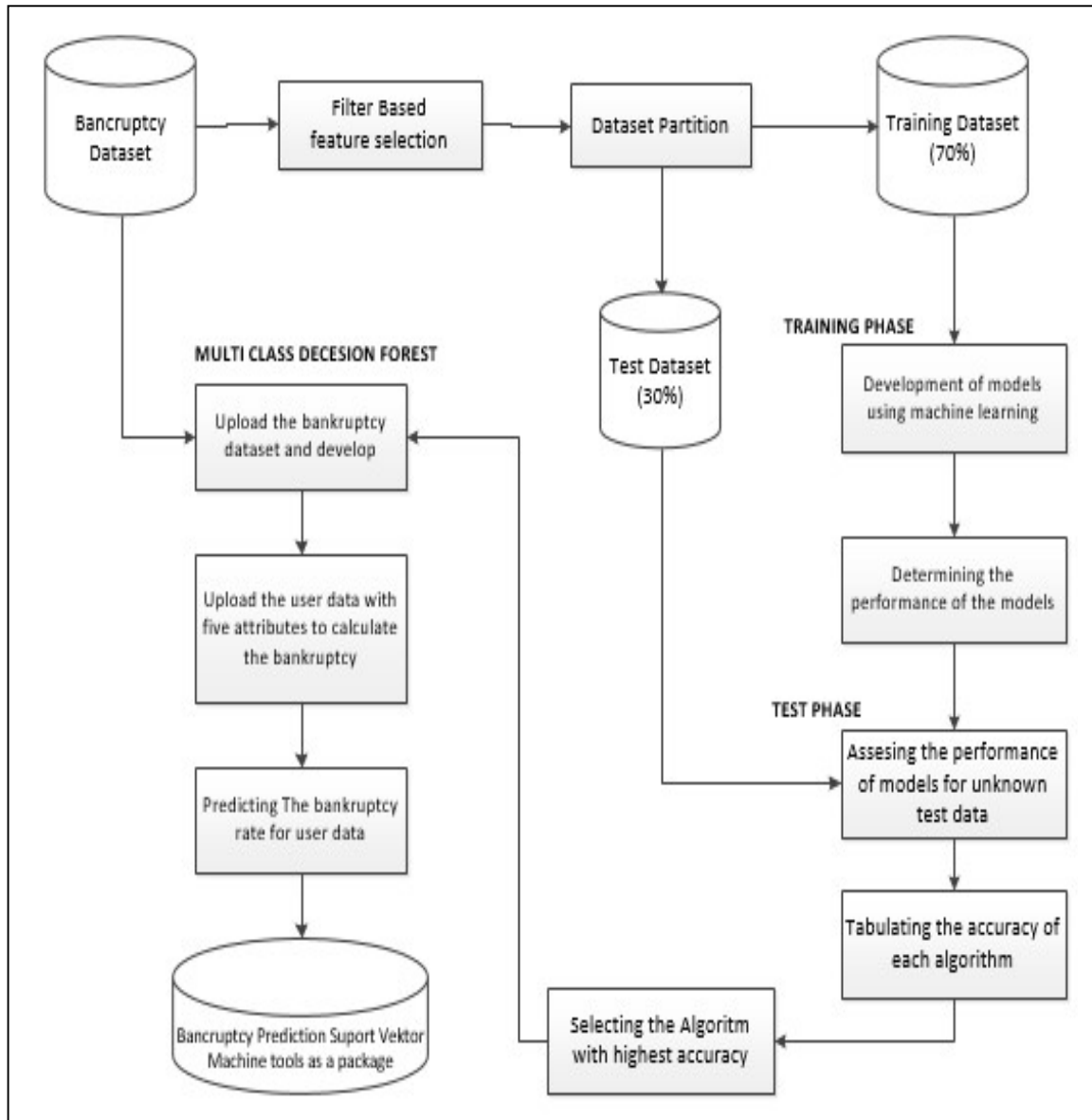
UCAPAN TERIMA KASIH

Terima kasih saya ucapkan kepada LPPM Universitas Maksoem yang telah memberikan kesempatan untuk melakukan penelitian, memberikan dukungan baik materil maupun non materil. Semoga Allah SWT memberikan balasan yang lebih baik.

DAFTAR PUSTAKA

- [1] W. Herdinytas and L. S. Almilia, 'Analisis rasio CAMEL terhadap prediksi kondisi bermasalah pada lembaga perbankan periode 2000-2002', *J. Akunt. dan Keuang.*, vol. 7, no. 2, pp. 131–147, 2006.
- [2] A. F. Irpanto, 'Analisis pengaruh rasio camel terhadap prediksi kondisi bermasalah pada lembaga perbankan', Universitas Negeri Malang, 2012.
- [3] D. E. Kurniawan, A. Saputra, P. Prasetyawan, and others, 'Perancangan Sistem Terintegrasi pada Aplikasi Siklus Akuntansi dengan Evaluasi Technology Acceptance Model (TAM)', *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 315–321, 2018.
- [4] F. Pedregosa *et al.*, 'Scikit-learn: Machine learning in Python', *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [5] F. Barboza, H. Kimura, and E. Altman, 'Machine learning models and bankruptcy prediction', *Expert Syst. Appl.*, vol. 83, pp. 405–417, 2017.
- [6] L. Zhou, K. K. Lai, and J. Yen, 'Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation', *Int. J. Syst. Sci.*, vol. 45, no. 3, pp. 241–253, 2014.
- [7] R. P. Hauser and D. Booth, 'Predicting bankruptcy with robust logistic regression', *J. Data Sci.*, vol. 9, no. 4, pp. 565–584, 2011.
- [8] Y. Hua, J. Guo, and H. Zhao, 'Deep belief networks and deep learning', in *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things*, 2015, pp. 1–4.
- [9] A. SETIAWAN MALAKA, 'MODEL PREDIKSI KEPAILITAN BANK UMUM DI INDONESIA MENGGUNAKAN ALGORITMA BACKPROPAGATION', *J. Ilmu Manaj.*, vol. 2, no. 4, 2015.
- [10] A. S. Malaka, 'Hartojo, "Model Prediksi Kepailitan Bank Umum Di Indonesia Menggunakan Algoritma Backpropagation,"' *J. Ilmu Manaj.*, vol. 2, no. 4, pp. 1714–1724, 2014.
- [11] J. Yao, N. Teng, H.-L. Poh, and C. L. Tan, 'Forecasting and analysis of marketing data using neural networks', *J. Inf. Sci. Eng.*, vol. 14, no. 4, pp. 843–862, 1998.
- [12] Y. Mishina, R. Murata, Y. Yamauchi, T. Yamashita, and H. Fujiyoshi, 'Boosted random forest', *IEICE Trans. Inf. Syst.*, vol. 98, no. 9, pp. 1630–1636, 2015.
- [13] G. Wang, J. Ma, and S. Yang, 'An improved boosting based on feature selection for corporate bankruptcy prediction', *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2353–2361, 2014.
- [14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 'Dropout: a simple way to prevent neural networks from overfitting', *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [15] P. Prasetyawan, I. Ahmad, R. I. Borman, Ardiansyah, Y. A. Pahlevi, and D. E. Kurniawan, 'Classification of the Period Undergraduate Study Using Back-propagation Neural Network', in *2018 International Conference on Applied Engineering (ICAE)*, 2018, pp. 1–5.
- [16] A. Dzikri and D. E. Kurniawan, 'Hand Gesture Recognition for Game 3D Object Using The Leap Motion Controller with Backpropagation Method', in *2018 International Conference on Applied Engineering (ICAE)*, 2018, pp. 1–5.

Lampiran



Gambar 4. Metode Penelitian (Tahapan Penerapan)