

## Original Article

e-ISSN: 2774-2016 - <https://journal.itera.ac.id/index.php/indojam/>  
p-ISSN: 2774-2067

Received 11th March 2023  
Accepted 21st March 2023  
Published 31st July 2023

Open Access

DOI:

10.35472/indojam.v3i1.1284

## Prediksi *Financial Distress* pada Salah Satu Bank Konvensional Menggunakan *Machine Learning*

Fuji Lestari <sup>\*a</sup>

<sup>a</sup> Program Studi Sains Aktuaria, Jurusan Sains, Intitut Teknologi Sumatera

\*Corresponding E-mail: [fuji.lestari@at.itera.ac.id](mailto:fuji.lestari@at.itera.ac.id)

**Abstract:** Financial distress is when a company experiences a shortage or insufficient funds to run the company. Prediction of financial distress is needed to prevent bankruptcy. In this study, financial distress predictions were made based on financial ratios obtained from monthly financial reports from a bank convention, after which the proportion that had the most influence on financial distress was determined. The models used in this study are several machine learning models, namely, Logistic Regression, Support Vector Machine, and Random Forest. Based on the analysis results, the best model for predicting financial pressure is the Random Forest Model, with an accuracy of 96.77%. Based on the best model obtained, namely the Random Forest, it can be determined that the ratio that is very influential on financial distress is the ratio of Total Asset Turnover.

**Keywords:** *Financial Distress, Machine Learning, Random Forest, Prediction*

**Abstrak:** *Financial distress* merupakan suatu keadaan perusahaan saat mengalami kekurangan atau ketidakcukupan dana dalam menjalankan perusahaan. Prediksi *financial distress* sangat untuk dibutuhkan untuk mencegah adanya kebangkrutan. Pada penelitian ini, dilakukan prediksi *finansial distress* berdasarkan rasio-rasio keuangan yang diperoleh dari laporan keuangan bulanan dari sebuah bank konvensional, setelah itu ditentukan rasio yang paling berpengaruh terhadap *financial distress*. Model yang digunakan dalam penelitian ini adalah beberapa model *machine learning* yaitu, *Logistic Regression, Support Vector Machine, dan Random Forest*. Berdasarkan hasil analisis, model terbaik dalam memprediksi *financial ditress* adalah Model *Random Forest* dengan akurasi sebesar 96,77%. Berdasarkan model terbaik yang diperoleh, yaitu *Random Forest*, dapat ditentukan rasio yang sangat berpengaruh pada *financial distress* adalah rasio *Total Asset Turnover*.

**Kata Kunci:** *Financial Distress, Machine Learning, Random Forest, Prediksi*

### Pendahuluan

Perekonomian merupakan salah satu komponen penting dalam melihat pertumbuhan suatu negara. Pertumbuhan perekonomian yang baik mengimplikasikan pertumbuhan negara yang baik pula. Namun, perekonomian negara saat ini sedang tidak dalam kondisi baik. Hal ini terlihat dari kondisi proyeksi perekonomian Indonesia yang sedang mengalami trend yang turun. Berdasarkan proyeksi dari *World Economic Outlook IMF* pada tahun 2022, ekonomi dunia mengalami pertumbuhan ekonomi sebesar 3,2% dan tahun berikutnya pertumbuhan ekonomi dunia juga diperkirakan akan semakin

melemah di angka 2,7%. Hal ini dipengaruhi oleh inflasi yang cukup tinggi sehingga pertumbuhan ekonomi dunia mengalami penurunan. Akibatnya, keadaan ini dapat merepresentasikan kondisi keuangan global yang tidak baik.

Keadaan ekonomi tersebut mempengaruhi banyak industri yang mendukung dalam meningkatkan perekonomian negara. Salah satunya adalah industri perbankan. Perbankan mempunyai peranan penting dalam perekonomian sebagai lembaga intermediasi yang menyalurkan dana masyarakat ke dalam investasi aset produktif yang akan mendorong produktivitas sektor riil, akumulasi kapital, dan pertumbuhan output agregat [1]. Berdasarkan data

## Original Article

Badan Pusat Statistik (BPS), sektor jasa keuangan dan asuransi berkontribusi 4,34% terhadap Produk Domestik Bruto nasional yang nilai totalnya mencapai Rp16,97 kuadriliun pada 2021 [2].

Kondisi ekonomi yang sedang melemah akan berdampak langsung pada industri perbankan. Oleh karena itu perlu dilakukan langkah pencegahan dengan terlebih dahulu menganalisis sistem yang ada di perusahaan sehingga bisa menanggulangi keadaan ekonomi global yang diproyeksikan akan melemah. Langkah pencegahan yang dapat dilakukan adalah dengan menganalisis faktor-faktor rasio keuangan yang mempengaruhi kebangkrutan dari sebuah perusahaan. *Financial distress* merupakan suatu keadaan perusahaan saat mengalami kekurangan atau ketidakcukupan dana dalam menjalankan perusahaan.

Analisis rasio keuangan tersebut perlu dilakukan oleh perusahaan sehingga perusahaan bisa menentukan langkah yang akan diambil selanjutnya. Rasio-rasio keuangan tersebut dapat diukur dari laporan keuangan. Laporan keuangan merupakan alat informasi yang menunjukkan kondisi kesehatan keuangan perbankan. Oleh karena itu, pada penelitian ini akan dilakukan perhitungan rasio keuangan yang diperoleh dari laporan bulanan salah satu bank konvensional. Dari rasio-rasio keuangan tersebut akan diprediksi *financial distress* dengan beberapa model *machine learning* dan akan ditentukan rasio-rasio yang berpengaruh pada *financial distress* menggunakan model terbaik yang telah diperoleh.

## Metode

Analisis rasio keuangan terhadap *financial distress* dapat dilakukan menggunakan beberapa metode. Dalam penelitian ini akan digunakan beberapa model *machine learning*, yaitu *Logistic Regression*, *Random Forest*, dan *Support Vector Machine*.

### 1. Logistic Regression

*Logistic Regression* adalah regresi yang digunakan untuk penelitian yang mempunyai variabel respon biner [3]. Oleh karena itu, model ini cocok digunakan dalam penelitian ini mengingat data *financial distress*

yang dipakai adalah biner. Persamaan regresi logistic dapat dilihat sebagai berikut.

$$P_n(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}} \quad (1)$$

Keuntungan model regresi logistik adalah hasil dari model ini mudah ditafsirkan dan dipahami. Model ini merupakan salah satu model yang sering digunakan dalam praktik di lembaga keuangan.

### 2. Support Vector Machine

*Support Vector Machine* (SVM) merupakan salah satu metode *machine learning* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada input space [4]. Metode ini menggunakan hipotesis berupa fungsi – fungsi linier dalam sebuah ruang fitur yang berdimensi tinggi, dengan mengimplementasikan pembelajaran bias yang berasal dari teori pembelajaran statistik. Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan. Proses klasifikasi akan menggunakan persamaan berikut.

$$y = \begin{cases} +1, & \text{jika } b + \alpha^T x \geq +1 \\ -1, & \text{jika } b + \alpha^T x \leq -1 \end{cases} \quad (2)$$

dengan  $b$  merupakan bias.

Fungsi kernel yang digunakan adalah sebagai berikut.

$$K(x, x_i) = \exp(-\gamma \|x - y\|^2)$$

Kekuatan SVM adalah menghindari *overfitting* dengan sampel kecil dan kurang sensitif terhadap distribusi yang tidak seimbang.

### 3. Random Forest

Dalam model ini, banyak *decision tree* dibangun menggunakan himpunan bagian dari fitur yang dipilih secara acak [3]. Bagian sampel dan fitur dipilih secara acak untuk memastikan keragaman dari klasifikasinya. Kemudian, *random forest* dibangun untuk beberapa subset yang sama dengan yang dihasilkan pada jumlah pohon klasifikasinya. Pada model ini menghasilkan

hasil yang lebih akurat dan cenderung menghindari *overfitting* dari model.

#### 4. Evaluasi Model

Dalam mengevaluasi model yang dipakai dalam penelitian ini, maka digunakan *performance metrics*.

##### 4.1. Akurasi

Akurasi menjelaskan proporsi klasifikasi yang benar dalam data penilaian ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Keterangan:

*True Positive (TP)*: Jumlah data testing yang diklasifikasikan model dengan kelas sebenarnya

*False Positive (FP)*: Jumlah data testing pada kolom yang sesuai kelasnya, namun bukan termasuk *TP*

*False Negative (FN)*: Jumlah data testing pada baris yang sesuai kelasnya, namun bukan termasuk *TP*

*True Negative (TN)*: Jumlah data testing pada setiap kolom dan baris namun tidak termasuk kolom dan baris kelas tersebut.

##### 4.2. Presisi (*Precision*)

Presisi menjelaskan proporsi *true positive* di antara prediksi positif.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

##### 4.3. Sensitivitas (*Recall*)

Sensitivitas menjelaskan proporsi positif yang diprediksi dengan benar.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

##### 4.4. $F1_{score}$

$F1_{score}$  menjelaskan tentang rata-rata harmonis dari *precision* dan *recall*.

$$F1_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

##### 4.5. Receiver Operating Characteristic (ROC)

ROC memplot hubungan antara tingkat *true positive* dengan tingkat *false positive* saat ambang keputusan berubah.

##### 4.6. Area di bawah kurva (AUC)

AUC bisa ditafsirkan sebagai proporsi sampel yang diklasifikasikan dengan benar. AUC juga membantu menyediakan kriteria untuk mengevaluasi dan membandingkan model.

## Hasil dan Diskusi

Data yang digunakan dalam penelitian ini adalah data laporan keuangan bulanan dari salah satu bank konvensional dari Januari 2015 sampai dengan Desember 2021. Variabel dependen dalam penelitian ini adalah *financial distress*. Sedangkan variabel independent dari penelitian ini adalah rasio keuangan yang dihitung berdasarkan laporan keuangan bulanan yang ditampilkan pada Tabel 1.

Tabel 1. Variabel Independen

Variabel	Definisi	Indikator
$X_1$	Rasio lancar ( <i>Current Ratio</i> )	$\frac{\text{Aset lancar}}{\text{hutang lancar}}$
$X_2$	<i>Return on Asset</i> (ROA)	$\frac{\text{Laba bersih sebelum pajak}}{\text{total aktiva}}$
$X_3$	<i>Return on Equity</i> (ROE)	$\frac{\text{Laba bersih setelah pajak}}{\text{total modal}}$
$X_4$	<i>Net Profit Margin</i> (NPM)	$\frac{\text{Laba bersih sebelum pajak}}{\text{penjualan bersih}}$
$X_5$	<i>Debt to Equity Ratio</i> (DER)	$\frac{\text{Total hutang}}{\text{modal}}$
$X_6$	<i>Debt to Total Asset</i> (DTR)	$\frac{\text{Total hutang}}{\text{Total aktiva}}$
$X_7$	<i>Total Asset Turnover</i>	$\frac{\text{Penjualan Bersih}}{\text{Total aktiva}}$

## Original Article

Tabel 2 merupakan tabel yang menjelaskan statistik deskriptif dari data variabel independen dalam penelitian ini.

**Tabel 2.** Statistika Deskriptif

Model	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
Mean	1,37	0,413	0,434	22,74	6,025	0,857	0,427
Max	1,6	0,984	0,997	1,848	7,116	0,877	0,993
Min	1,31	0,115	0,111	1,197	5,296	0,835	0,111
Std.Dev	0,04	0,272	0,259	61,35	0,425	0,009	0,244
Skew	2,2	0,502	0,498	8,846	0,494	0,069	0,622
Kurt	7,88	-1,33	-0,69	80,02	-0,47	-0,15	-0,48

Pada penelitian ini, akan dilihat perbedaan dari model *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* dalam hal mengevaluasi *financial distress* pada bank konvensional yang dilihat dari laporan keuangan bulanan. Pada penelitian ini, akan dilihat kinerja dari masing masing model dengan melihat akurasi dan grafik AUC-ROC. Hasil akurasi dari masing-masing model dapat dilihat pada Tabel 3.

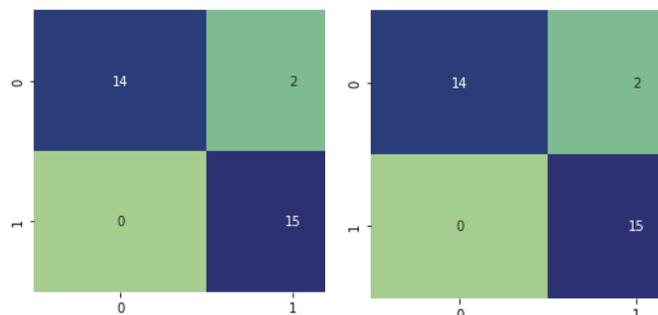
**Tabel 3.** Akurasi Data Train dan Test

Model	Akurasi Data Train	Akurasi Data Test
<i>Logistic Regression</i>	95,04%	93,55%
<i>Support Vector Machine</i>	95,04%	93,55%
<i>Random Forest</i>	93,55%	96,77%

Berdasarkan akurasi dari tiap model, akurasi dari model *Logistic regression* dan *Support Vector Machine* sebesar 93,55%, sedangkan model *Random Forest* sebesar 96,77%. Dari hasil tersebut dapat dilihat bahwa model *Random Forest* memiliki akurasi yang tinggi dalam memprediksi *financial distress* dari salah satu bank konvensional. Hal ini sesuai dengan literatur yang mengatakan bahwa model *Random Forest* akan menghasilkan kinerja prediksi yang baik.

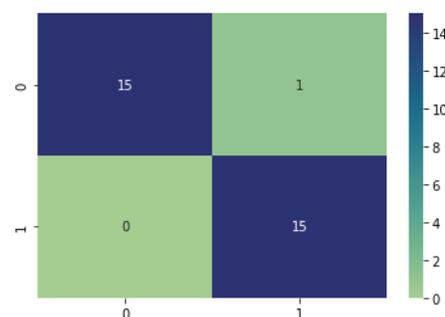
Berdasarkan *matrices confusion* pada Gambar 1-3, dapat dilihat bahwa nilai TP tertinggi menggunakan model *Random Forest* dibandingkan model *Logistic Regression* dan *Support Vector Machine*, yaitu 15 data diprediksi dengan benar. Hal ini juga mendukung hasil akurasi

yang menyatakan model *Random Forest* lebih baik dari model lainnya.



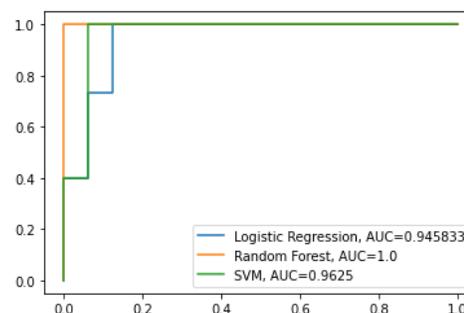
**Gambar 1.** Matriks Confusion LR

**Gambar 2.** Matriks Confusion SVM



**Gambar 3.** Matriks Confusion RF

Selanjutnya akan dilihat grafik ROC-AUC dari setiap model sebagai berikut.

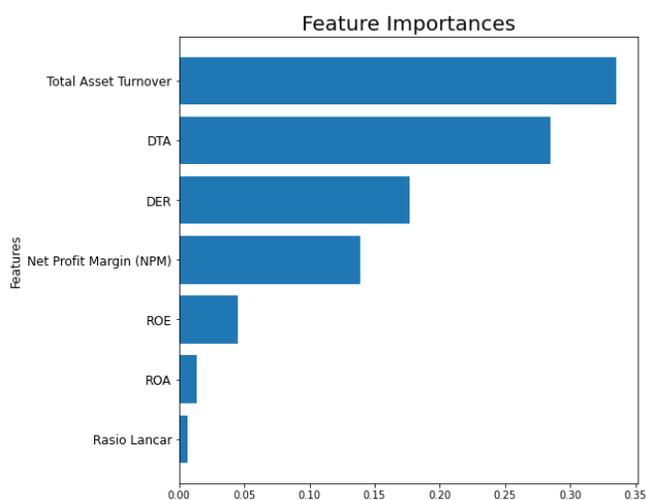


**Gambar 4.** Plot ROC

AUC adalah probabilitas bahwa pengklasifikasi memberi peringkat sampel positif yang dipilih secara

acak lebih tinggi daripada sampel negatif yang dipilih secara acak. Kurva yang mendekati pojok kiri atas bagan dapat diklasifikasikan menjadi model yang baik, karena tingkat *true positive* nya mendekati 100% sedangkan tingkat *false positive* nya 0%. Pada Gambar 4. Dapat terlihat bahwa AUC dari model *Logistic Regression* diperoleh sebesar 0,945833, model *SVM* sebesar 0,9625, sedangkan model *Random Forest* sebesar 1,0. Berdasarkan hasil tersebut dapat dilihat bahwa nilai AUC dari model *Random Forest* adalah yang terbaik dibandingkan dengan model lainnya. Sehingga model *Random Forest* dapat dikatakan model yang baik dalam memprediksi *financial distress*.

Setelah ditentukan model terbaik dalam memprediksi *financial distress*, maka selanjutnya akan dicari variabel yang sangat berpengaruh pada *financial distress* yang dapat dilihat pada Gambar 5 berikut.



Gambar 5. Variabel Terbaik

Berdasarkan hasil tersebut, dapat disimpulkan bahwa variabel yang sangat berpengaruh pada *financial distress* menggunakan model *Random Forest* adalah rasio *Total Asset Turnover*.

## Kesimpulan

Berdasarkan hasil analisis di atas dapat disimpulkan bahwa

1. Model *Random Forest* merupakan model terbaik dalam memprediksi *Financial Distress* dibandingkan dengan model *Logistic Regression* dan *Support Vector Machine* dengan akurasi sebesar 96,77%.
2. Variabel yang paling berpengaruh pada *financial distress* menggunakan model *Random Forest* adalah rasio *Total Asset Turnover*.

## Konflik Kepentingan

Penulis menyatakan bahwa artikel ini tidak memiliki konflik kepentingan tentang publikasi.

## Referensi

- [1] V. R. Bencivenga dan B.D. Smith, "Financial intermediation and endogenous growth", *The Review of Economic Studies*, Vol. 58: 195-209, 1991.
- [2] Badan Pusat Statistik (2020). *Pertumbuhan Ekonomi Indonesia Tahun 2020*. Jakarta.
- [3] K. L. Tran dkk, "Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam," *MDPI*, Vol. 7: 160, 2022.
- [4] I. M. Parapat, M. T. Furqon, dan Sutrisna. "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 10: 3163-3169, 2018.
- [5] (IMF), I.N. (2022). *World Economic Outlook (International Monetary Fund)*. Washington: International Monetary Fund.