Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) p-ISSN: 2355-7699
Vol. 12, No. 5, Oktober 2025, hlm. 1209-1220 e-ISSN: 2528-6579

PENERAPAN *RANDOM OVERSAMPLING* DAN *PRINCIPAL COMPONENT ANALYSIS* UNTUK MENINGKATKAN AKURASI PREDIKSI KEBANGKRUTAN PERUSAHAAN DI INDONESIA DENGAN MODEL *MACHINE LEARNING*

Zainil Abidin[*1], Tri Suratno[2] , Mutia Fadhila Putri[3]

[1,2,3] Universitas Jambi, Kabupaten Muaro Jambi
Email: [1]zainil.abidin@unja.ac.id, [2]tri@unja.ac.id, [3]mutia.fadhila@unja.ac.id
[*]Penulis Korespondensi

(Naskah masuk: 17 Maret 2025, diterima untuk diterbitkan: 30 April 2025)

Abstrak

Prediksi kebangkrutan menjadi penting untuk memberikan peringatan dini bagi manajemen dan pemangku kepentingan agar dapat mengambil tindakan preventif. Penelitian ini menguji penerapan metode *Random Oversampling* dan *Principal Component Analysis* (PCA) dalam model *machine learning* untuk meningkatkan akurasi prediksi kebangkrutan perusahaan. Penelitian ini menggunakan dua dataset yaitu data *Taiwanese Bankruptcy Prediction* dari *UCI Machine Learning Repository* sebanyak 6.891 data dan data primer berupa data kebangkrutan perusahaan Indonesia dari Bursa Efek Indonesia (BEI) dari tahun 2021-2023 sebanyak 2.703 data. Total keseluruhan dataset yang digunakan sebanyak 9.594 data. Empat algoritma klasifikasi—KNN, Naïve Bayes, SVM, dan Decision Tree—diuji sebelum dan sesudah penerapan metode tersebut. Hasil menunjukkan bahwa kombinasi PCA dan *Random Oversampling* meningkatkan *recall* kelas minoritas (kebangkrutan) secara signifikan. SVM menjadi algoritma terbaik dengan *precision* 0,86, *recall* 0,76, dan *F1-score* 0,80, sementara *Decision Tree* mengalami *overfitting* setelah *oversampling*. PCA berhasil mereduksi dimensi dataset hingga 98,87% varian tetap terjaga, dan *Random Oversampling* menyeimbangkan distribusi kelas.

Kata kunci: prediksi kebangkrutan, *machine learning*, *random oversampling*, PCA.

APPLICATION OF RANDOM OVERSAMPLING AND PRINCIPAL COMPONENT ANALYSIS TO ENHANCE THE ACCURACY OF BANKRUPTCY PREDICTION FOR COMPANIES IN INDONESIA USING MACHINE LEARNING MODELS

Abstract

Bankruptcy prediction is crucial for providing early warnings to management and stakeholders to take preventive actions. This study examines the application of Random Oversampling and Principal Component Analysis (PCA) in machine learning models to improve the accuracy of corporate bankruptcy prediction. The study uses two datasets: the Taiwanese Bankruptcy Prediction data from the UCI Machine Learning Repository (6,891 data points) and primary data on Indonesian company bankruptcies from the Indonesia Stock Exchange (IDX) for 2021–2023 (2,703 data points), totaling 9,594 data points. Four classification algorithms—K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM), and Decision Tree—were tested before and after applying these methods. The results show that the combination of PCA and Random Oversampling significantly improved the recall of the minority class (bankruptcy). SVM emerged as the best-performing algorithm with a precision of 0.86, recall of 0.76, and F1-score of 0.80, while the Decision Tree experienced overfitting after oversampling. PCA successfully reduced the dataset's dimensions while retaining 98.87% of the variance, and Random Oversampling balanced the class distribution.

Keywords: *bankruptcy prediction, machine learning, random oversampling, PCA*

1. PENDAHULUAN

Istilah kebangkrutan dapat didefinisikan sebagai kondisi dimana suatu perusahaan tidak mampu lagi menjalankan operasional yang umumnya disebabkan oleh masalah kesulitan keuangan (Rahayu and Usmansyah, 2021a). Kebangkrutan perusahaan akan berdampak kepada *stakeholder* seperti para investor, karyawan dan pihak lain yang memiliki keterkaian dengan dengan perusahaan. Kebangkrutan juga akan menyebabkan terjadinya efek domino seperti menganggu kestabilan pasar dan meningkatnya gelombang PHK (Diana and Hidayat, 2023). Permasalahan kesulitan keuangan yang menyebabkan terjadinya

kebangkrutan pada suatu perusahaan tidak terjadi dalam waktu yang cepat. Permasalahan ini bisa diprediksi dengan meneliti laporan keuangan yang dimiliki perusahaan. Prediksi kebangkrutan melalui analisis laporan keuangan (rasio keuangan) perusahaan merupakan langkah awal yang penting, agar jajaran manajemen perusahaan dan pihak yang berkempentingan bisa mencegah terjadinya kebangkrutan (Rahayu and Usmansyah, 2021a).

Dalam ilmu ekonomi, ada beberapa metode untuk memprediksi kebangkrutan berdasarkan rasio keuangannya. Model yang sering peneliti ekonom adalah *altman z-score*, metode *springate*, metode *zmijewski*, metode *foster*, dan metode *Grover* (Setiawan, 2021). Metode prediksi kebangkrutan tersebut menggunakan kombinasi rasio keuangan tertentu, sehingga hasil dari prediksi tidak bisa memberikan cerminan keseluruhan kondisi keuangan perusahaan. Alternatif lain yang bisa diimplementasikan untuk mengatasi permasalahan ini adalah dengan *machine learning* (Iparraguirre-Villanueva and Cabanillas-Carbonell, 2024). Pengimplementasian *machine learning* dalam memprediksi kebangkrutan mampu memperluas penggunaan rasio-rasio keuangan yang lebih banyak dan kompleks sehingga hasil dari prediksi menjadi lebih akurat (Aljawazneh et al., 2021b; Dasilas and Rigani, 2024).

Machine learning bekerja berdasarkan pola dari dataset yang dilatih dengan menggunakan algoritma pembelajaran tertentu. Pola ini terbentuk dari data historis rasio keuangan perusahaan yang bangkrut ataupun yang masih bertahan (tidak bangkrut) (Shetty, Musa and Brédart, 2022). Beberapa rasio yang digunakan adalah likuiditas, solvabilitas, profitabilitas, aktivitas, dan rasio ukuran perusahaan. Rasio ini akan menjadi *role model* sistem untuk mempelajari seperti apa pola rasio keuangan perusahaan yang diprediksikan bangkrut (Papík and Papíková, 2024). Model ini kemudian digunakan kembali untuk memprediksi kemungkinan kebangkrutan perusahaan lainnya. Membuat suatu model *machine learning* sangat perlu memperhatikan algoritma yang digunakan (Iparraguirre-Villanueva and Cabanillas-Carbonell, 2024).

Salah satu aspek krusial dalam mengimplementasikan *machine learning* untuk memprediksi kebangkrutan adalah algoritma pembentuk model prediksi (Galil, Hauptman and Rosenboim, 2023). Algoritma *machine learning* sangat penting dipilih karena setiap algoritma menghasilkan tingkat akurasi yang berbeda-beda (Park et al., 2021a). Sistem prediksi merupakan bagian dari algoritma *supervised learning* yang masuk kategori klasifikasi. Ada beberapa algoritma klasifikasi yang bisa digunakan pada sistem prediksi kebangkrutan yaitu *decision tree*, *K-nearest neighbors* (KNN), *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Random Forest* (Gurnani et al., 2021; Park et al., 2021b). Algoritma *machine learning* tersebut bekerja berdasarkan dataset yang diberikan, semakin tinggi kualitas dataset maka model prediksi yang dihasilkan juga akan bagus (Aljawazneh et al., 2021a). Namun, permasalahan yang sering dihadapi adalah dataset model *machine learning* yang tidak seimbang (*imbalanced dataset*) dan reduksi dimensi dataset yang tinggi (Byun et al., 2025).

Permasalahan yang sering dihadapi dalam membangun *machine learning* adalah permasalahan keseimbangan data atau dikenal dengan *imbalanced dataset (Das et al., 2020)*. Ketidakseimbangan ini dikarenakan jumlah pada dataset kelas mayoritas lebih (tidak bangkrut) lebih banyak dibandingkan kelas minoritas (bangkrut). Hal ini akan membuat model *machine learning* lebih bias pada kelas yang mayoritas dibandingkan dengan kelas minoritas(Pranavi et al., 2022). Jika masalah ini dibiarkan akan mengurangi performa model untuk memprediksi perusahaan yang berisiko bangkrut. Ketidakseimbangan dataset dapat diatas dengan menggunakan *random oversampling (Liu et al., 2022)*. Metode ini digunakan pada penelitian karena penggunaan metode yang sederhana dan mampu meningkatkan representasi kelas yang minoritas lebih baik tanpa menghilangkan data asli dari dataset (Tarawneh et al., 2022).

Random Oversampling dipilih untuk mengatasi ketidakseimbangan kelas dalam dataset. Dataset kebangkrutan bank sering kali memiliki proporsi yang tidak seimbang antara perusahaan yang bangkrut (kelas minoritas) dan yang tidak bangkrut (kelas mayoritas). Ketidakseimbangan ini dapat menyebabkan model *machine learning* cenderung bias terhadap kelas mayoritas, sehingga performa prediksi untuk kelas minoritas menjadi rendah (Pranavi et al., 2022). Dengan menerapkan random oversampling, distribusi kelas dalam dataset menjadi lebih seimbang, sehingga model dapat belajar pola dari kedua kelas secara lebih optimal (Binanto et al., 2024).

Proses membangun model prediksi kebangkrutan Perusahaan juga dipengaruhi oleh dimensi tinggi dari fitur kebangkrutan yang digunakan yaitu variabel rasio keuangan perusahaan. Prediksi kebangkrutan memiliki banyak variabel yang bisa digunakan seperti likuiditas, aktivitas, profitabilitas dan rasio keuangan lainnya (Lord et al., 2020; Rahayu and Usmansyah, 2021b). Semakin banyak variabel yang digunakan seharusnya memberikan peluang prediksi yang semakin baik. Namun, Dalam *machine learning* hal tersebut akan menyebabkan terjadinya *overfitting* dan meningkatkan kompleksitas komputasi model yang dibangun. Permasalahan *overfitting* pada model *machine learning* prediksi kebangkrutan akan menyebabkan model tidak bisa begitu akurat memprediksi data yang baru (Kaib et al., 2024). *Principal Component Analysis* (PCA) merupakan

metode yang banyak digunakan untuk menangani permasalahan ini (He and Zhang, 2020).

Metode PCA digunakan untuk mengurangi reduksi dimensi data dengan tetap mempertahankan informasi penting yang ada dalam dataset. PCA memiliki keunggulan dalam menangkap varian maksimum yang terdapat pada data rasio keuangan tanpa bergantung pada label kelas (Yaicharoen and Yamada, 2021). Metode *Principal Component Analysis* (PCA) dipilih untuk mereduksi dimensi dataset yang memiliki banyak variabel keuangan. Hal ini dilakukan untuk mengatasi masalah *multicollinearity* (kolinearitas antar variabel) yang sering ditemui dalam analisis data keuangan. Dengan PCA, dimensi dataset dapat dikurangi menjadi komponen utama yang mampu menjelaskan sebagian besar variasi data, sehingga model *machine learning* dapat bekerja lebih efisien tanpa kehilangan informasi penting (Yaicharoen and Yamada, 2021; Jasim et al., 2024).

Pada beberapa penelitian sebelumnya, telah membahas pengimplementasian algoritma *machine learning* guna memprediksi kebangkrutan Perusahaan. Seperti pada penelitian Quynh & Phoung (2020), Gurnani et al (2021), Aljawazneh et al (2021), Park et al (2012), Khalil et al (2022), dan Siswoyo, dkk (2022). Akan tetapi masih sedikit peneliti yang melakukan penelitian dengan mengkombinasikan metode *Oversampling* dan PCA pada algoritma *machine learning* khususnya dalam konteks prediksi kebangkrutan perusahaan Indonesia selama periode tertentu.

Pada penelitian ini, peneliti bertujuan mengkombinasikan metode *Oversampling* dan PCA pada algoritma *machine learning* untuk membandingkan akurasi algoritma memprediksi kebangkrutan perusahaan. Kombinasi PCA dan *random oversampling* dipilih untuk meningkatkan performa model prediksi kebangkrutan. PCA digunakan untuk mereduksi dimensi dataset dan menghilangkan redundansi antar variabel, sementara random oversampling digunakan untuk menyeimbangkan distribusi kelas dalam dataset. Dengan pendekatan ini, model machine learning dapat bekerja lebih efisien dan adil dalam memprediksi kemungkinan kebangkrutan Perusahaan.

2. TINJAUAN PUSTAKA

Istilah kebangkrutan dapat diartikan sebagai indikator Perusahaan yang tidak mampu mengelola dana yang dimiliki sehingga operasional perusahaan tidak bisa berjalan (Putri and Challen, 2021). Beberapa peneliti telah berusaha melakukan kajian guna memprediksi kebangkrutan sebuah perusahaan supaya masalah kebangkrutan tersebut bisa diantisipasi. Salah satu teknik yang dilakukan peneliti adalah dengan mengkaji rasio keuangan perusahaan (Masdiantini and Warasniasih, 2020). Rasio keuangan seperti likuiditas, profitabilitas, solvabilitas, dan rasio lainnya bisa memberikan gambaran yang jelas dan mendalam berkaitan dengan kondisi keuangan Perusahaan. Berlandaskan rasio keuangan para peneliti ekonomi, berusaha mengembangkan model prediksi berbasiskan statistik seperti model *altman z-score*, model *Ohlson*, dll (Lord et al., 2020).

Namun, beberapa tahun terakhir penelitian terkait prediksi kebangkrutan perusahaan juga menarik dilakukan dengan metode *machine learning*. Penelitian Quynh & Phoung (2020) yang menggunakan data penelitian Polish companies' bankruptcy dataset dari tahun 2007 sampai 2013. Penelitian ini dilakukan dengan mengkombinasikan antara algoritma *forest regression* dengan *Synthetic Minority Oversampling Technique* (SMOTE). Hasil penelitian menunjukan akurasi terbaik yang dihasilkan oleh kombinasi algoritma pada nilai AUC sebesar 99,78% dan nilai *F1 Score* sebesar 95,12%.

Penelitian yang dilakukan oleh Gurnani et al (2021) melakukan penelitian untuk melihat akurasi algoritma *random forest* dengan algoritma machine learning lainnya dalam melakukan prediksi kebangkrutan perusahaan. Penelitian ini menggunakan dataset dari *Taiwan Economic Journal* yang berisi data rasio keuangan dari tahun 1999-2009. Hasil dari penelitian Gurnani et al (2021) menunjukan algoritma *random forest* mendapatkan nilai sebesar 97,8% dan nilai *F1 Score* sebesar 97,7%. Nilai akurasi ini lebih tinggi dibandingkan dengan algoritma KNN sebesar 92,2%, algoritma SVM sebesar 62,2%, dan algoritma *decision tree* sebesar 95%.

Penelitian Aljawazneh et al (2021) melakukan perbandingan dari beberapa algoritma *machine learning* dalam memprediksi kebangkrutan perusahaan. Penelitian ini menggunakan dataset dari perusahaan Spanyol, dataset perusahaan Polandia dan perusahaan Taiwan. Hasil penelitian menunjukan akurasi algoritma *random forest* menghasilkan akurasi sebesar 99,34%, algoritma SVM akurasinya sebesar 98,86%, dan algoritma KNN sebesar 98,95%.

Penelitian lainnya dilakukan oleh Park et al (2021) menggunakan data keuangan yang dikumpulkan dari perusahaan Korea sepanjang periode tahun 2009 sampai 2015. Data ini berasal dari *Douzone Bizon ICT Group*. Hasil penelitian menunjukan algoritma *random forest* menghasilkan nilai akurasi sebesar 96,1%. Penelitian Khalil et al (2022) yang bertujuan membangun model machine learning untuk memprediksi kebangkrutan pada perusahaan asuransi Mesir. Dataset dikumpulkan dari 11 perusahaan asuransi Mesir sepanjang periode 1999 sampai 2019. Hasil penelitian menunjukan model machine learning dengan menggunakan algoritma CatBoost memiliki akurasi sebesar 96,46%. Hasil akurasi model ini lebih tinggi dibandingkan nilai akurasi dengan algoritma lain seperti Algoritma Bagging sebesar 96,19%,

algoritma *random forest* sebesar 95,75%, dan akurasi algoritma *decision tree* sebesar 91,58%.

Pada penelitian khusus dataset perusahaan di Indonesia sudah pernah dilakukan beberapa peneliti sebelumnya. Penelitian Siswoyo, dkk (2022) menggunakan dataset history dataset perusahaan perbankan di Indonesia yang dikumpulkan untuk periode 2010 sampai 2016. Hasil penelitian menunjukan algoritma algoritma BELM menghasilkan akurasi sebesar 97%, algoritma *random forest* sebesar 90%, algoritma LRC dan SVM masing-masing akurasinya sebesar 81%. **Tabel 1** menunjukan ringkasan hasil dari penelitian terdahulu untuk penelitian prediksi kebangkrutan perusahaan dengan model machine learning.

Tabel 1. Penelitian Terdahulu

Peneliti	Dataset	Metode	Akurasi
Quynh & Phoung (2020)	Polish companies' bankruptcy dataset dari tahun 2007 sampai 2013	Forest Regression + SMOTE	FR = 99,78%
Gurnani et al (2021)	Taiwan Economic Journal yang berisi data dari tahun 1999-2009	Random Forest	RF = 97,8%
Aljawazneh et al (2021)	perusahaan Spanyol, dataset perusahaan Polandia dan perusahaan Taiwan	Random Forest, SVM, dan KNN	RF = 99,34% SVM = 98,86% KNN = 98,95%
Park et al (2021)	perusahaan Korea sepanjang periode tahun 2009 sampai 2015	Random Forest	RF = 96,1%
Khalil et al (2022)	perusahaan asuransi Mesir sepanjang periode 1999 sampai 2019	CatBoost, Bagging, Random forest, Decision Tree	CatBoost = 96,46% Bagging = 96,19& Random Forest = 95,75% Decision Tree = 91,58%
Siswoyo, dkk (2022)	Perusahaan perbankan di Indonesia periode 2010-2016	BELM, Random Forest, LRC, SVM	BELM = 97% Random Forest = 90% LRC = 81% SVM = 81%

3. MOTEODE PENELITIAN

2.1 Data Penelitian

Data penelitian yang digunakan untuk membangun model prediksi merupakan data sekunder yang diambil dari UCI Machine Learning Repository (Taiwanese Bankruptcy Prediction). Data ini diambil dari laporan keuangan perusahaan yang terdaftar di Taiwan Stock Exchange selama periode 1999 hingga 2009. Terdapat dua kelas utama yaitu perusahaan bangkrut dan perusahaan tidak bangkrut. Dataset ini berjumlah sebanyak 6819 data rasio keuangan perusahaan dan terdapat 95 variabel seperti rasio ROA, Current Ratio, Quick ratio, dan lain-lain. Variabel dalam dataset penelitian disesuaikan kembali sesuai dengan variabel teori kebangkrutan yang diteliti.

Selain dari data sekunder, peneliti menggunakan data primer yang berupa data keuangan perusahaan-perusahaan yang ada di Indonesia. Pengambilan dataset ini bertujuan memberikan kekhasan dataset khususnya perusahaan Indonesia. Data ini dikumpulkan dari semua sektor perusahaan indonesia yang terdaftar di Bursa Efek Indonesia (BEI) atau Indonesia *Stock Exchange* selama pada tahun 2023 yaitu sebanyak 901 perusahaan. Dari data tersebut, peneliti melakukan pengambilan rasio keuangan berdasarkan laporan keuangan perusahaan selama 3 tahun yaitu periode 2021-2023. Data primer yang didapatkan adalah sebanyak 2703 data rasio keuangan.

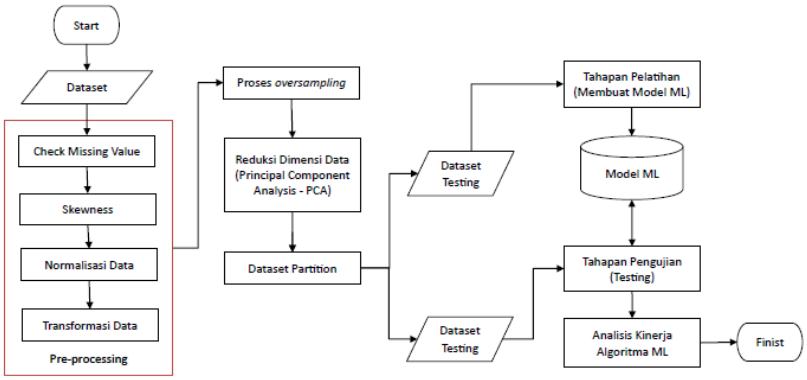
Jumlah keseluruhan data penelitian ini adalah sebanyak 9522 data rasio keuangan. Pengkombinasian dataset penelitian dari dataset perusahaan Taiwan dan perusahaan Indonesia bertujuan untuk membuat model prediksi yang bisa beradaptasi dengan konteks geografis maupun ekonomi.Hal ini diharapkan mampu menciptakan model prediksi kebangkrutan yang lebih general. **Tabel 2** menunjukan jumlah dataset yang digunakan pada penelitian ini.

Tabel 2. Dataset penelitian

No	Dataset	Bangkrut	Tidak Bangkrut	Jumlah
1	Dataset Taiwanese Bankruptcy Prediction periode 1999-2009	220	6.599	6.891
2	Dataset Perusahaan Indonesia periode 2021-2023 (*primer dataset*)	163	2540	2.703
	Total	**383**	**9.139**	**9.522**

2.2 Proses Annotasi Data

Pada dataset untuk perusahaan indonesia, belum ditentukan kelas atribut (bangkrut atau tidak). Oleh karena itu, diperlukan proses pelabelan kelas atribut (annotasi data) sebelum digunakan sebagai dataset pembuatan model. Proses Annotasi dataset perusahaan indonesia dilakukan dengan menggunakan model perhitungan *Ohlson O-Score (Seto, 2022)*. Pertimbangan penggunaan *Ohlson O-Score* dalam proses annotasi data antara lain model ini bersifat fleksibel untuk digunakan di berbagai industry (Widiasmara and Rahayu, 2019; Pamungkas, 2023). Model *Ohlson O-Score* bekerja berdasarkan probabilitas dengan melakukan perhitungan yang berdasarkan rasio likuiditas, profitabilitas, *leverage*, dan ukuran perusahaan (Seto, 2022).



Gambar 1. Model Machine Learning Prediksi Kebangkrutan perusahaan

2.3 Variabel Penelitian

Model pembelajaran prediksi kebangkrutan perusahaan dengan metode *machine learning* bekerja berdasarkan variabel yang digunakan (Chen, 2023). Variabel ini sebagai dasar penentuan kelas atribut. Kelas atribut atau label kelas terdiri dari dua yaitu bangkrut dan tidak bangkrut. Adapun variable rasio keuangan perusahaan ditunjukan pada **Tabel 3**.

Tabel 3. Rasio dan Variabel Penelitian

Rasio Keuangan	Variabel
Rasio Likuiditas	*Current Ratio*
	Quick Ratio
Rasio Aktivitas	*Inventory Turnover Ratio*
	Day Sales Outstanding
	Fixed Assets Turnover Ratio
	Total Assets Turnover
Rasio Leverage	*Debt to Total Assets atau Debt Ratio*
	Debt to Equity Ratio
	Times Interest Earned
Rasio Profitabilitas	*Gross profit margin*
	Net profit margin
	Operating Profit Margin
	Basic Earning Power (BEP)
	Return on Common Equity (ROE)
	Return on Assets (ROA)
Rasio Nilai Pasar	*Earning per Share*
	Price Earning Ratio (PER)
	Book Value Per Share (BVS)

2.4 Perancangan Model

Pada perancangan model *machine learning* prediksi kebangkrutan perusahaan terdiri dari *pre-processing data*, proses *random oversampling*, reduksi dimensi data dengan PCA, *dataset partition*, tahapan pelatihan (*training phase*), tahapan pengujian (*testing phase*), dan analisis kinerja algoritma. **Gambar 1** menunjukan tahapan Dalam membuat model machine learning prediksi kebangkrutan Perusahaan.

1. Pre-Processing

Pada tahapan pre-processing terdiri dari tahapan cek *missing values*, *skewness*, normalisai data dan transformasi data. Proses *missing values* bertujuan untuk mengecek apakah terdapat data yang kosong atau tidak. Jika proses pengecekan data terdapat data kosong, maka data akan diisi dengan nilai *median*.

Tahapan *pre-processing* selanjutnya adalah melihat kemiringan dataset atau *skewness*, yaitu ketidaksimetrisan atau kemiringan dari sebuah distribusi. Sebuah dataset yang mengalami kemiringan akan menyebabkan distribusi data berkonsentrasi hanya pada satu sisi. Dalam machine learning kemiringan data yang terlalu besar akan mempengaruhi akurasi model yang dihasilkan nantinya. Guna mengatasi permasalahan distribusi data yang terlalu miring bisa menggunakan metode transformasi data.

Tahapan selanjutnya adalah normalisasi atau standardisasi, yaitu memastikan bahwa titik dataset memiliki skala yang seimbang. Hal ini sangat penting Dalam *machine learning*, terutama ketika menggunakan pemodelan yang *sensitive* terhadap perbedaan skala fitur seperti pada algoritma SVM dan KNN). Proses standirdisasi menggunakan metode standarscaler, dimana mengubah data sehingga rata-rata (*mean*) fitur menjadi nol (berpusat pada nol) dan simpangan baku (standar deviasi) menjadi satu.

Tahapan terakhir pada *pre-processing* adalah proses transformasi. Tahapan Transformasi data merupakan bagian yang penting dalam membangun model *machine learning*. Transformasi data berguna untuk memastikan dataset model kompatibel dengan model *machine learning* yang dibuat sehingga mampu mengurangi masalah *outlier* serta meningkatkan akurasi model. Pre-processing dataset pada tahap transformasi data menggunakan metode *log transformation*. Metode ini adalah metode trasnformasi data yang mengantikan (*replace*) setiap variabel x dengan sebuah $\log(x)$.

2. Imbalanced dataset

Dataset pada *machine learning* terkadang memiliki jumlah kelas atribut atribut yang tidak sama, atau disebut dengan *imbalanced dataset*. Salah satu Teknik mengatasi *imbalanced dataset* adalah Metode random *Oversampling (Tarawneh et al., 2022)*. Cara kerja *Oversampling* dengan melakukan augmentasi data yang tidak seimbang pada kelas minoritas. Dimana Teknik ini akan

meningkatkan jumlah sampel dataset di kelas minoritas.

Cara kerja dari Teknik random *Oversampling* melakukan imbalanced dataset ditunjukan pada ***Pseudo-code*** **algoritma 1** berikut ini.

Algorithm 1. Metode Random *Oversampling* untuk Menyeimbangkan Dataset Imbalanced

1. **Input:** Dataset tidak seimbang D, himpunan fitur X, variabel target Y
2. **Proses:**
3. Identifikasi kelas mayoritas C_{maj} dan kelas minoritas C_{min}
4. Hitung jumlah sampel dalam C_{maj} dan C_{min}
5. Hitung selisih jumlah sampel: $\delta=|C_{maj}|-|C_{min}|$
6. Selama $\delta>0$:
7. Pilih secara acak satu sampel dari C_{min} dengan pengembalian (*replacement*)
8. Tambahkan sampel yang dipilih ke C_{min}
9. Perbarui nilai δ
10. Gabungkan C_{maj} dan C_{min} yang telah diperbarui menjadi dataset seimbang D`
11. **Output:** Dataset seimbang D′ dengan jumlah sampel yang sama untuk setiap kelas

3. Reduksi Dimensi Data

Reduksi dimensi data merupakan proses mengurangi dan mempertahankan jumlah variabel penting (*feature*) pada dataset agar model *machine learning* bisa lebih efisien. Selain itu, reduksi data juga bertujuan untuk mempercepat komputasi model dan mengatasi *overfitting (Adebusola et al., 2025)*. Penelitian ini menggunakan *Principal Component Analysis* (PCA), yaitu sebuah *dimensionality reduction* yang mengubah fitur dataset yang besar menjadi kumpulan fitur kecil tapi masih mempertahankan fitur yang signifikan (Xie et al., 2024). Dalam penelitian ini, sebelum melakukan proses reduksi terlebih dahulu dilakukan proses normalisasi. Tujuannya agar fitur yang terdapat pada dataset memiliki nilai *mean* nol dan standar deviasi 1. Sehingga dataset yang memiliki jumlah yang besar tidak mendominsi dataset berjumlah kecil. Jika ini terjadi, maka hasil dari proses PCA tidak akurat (Yaicharoen and Yamada, 2021).

Dataset hasil normalisasi akan dihitung matriks kovarians dengan menggunakan rumus berikut ini (Xue et al., 2024).

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1} \quad (1)$$

Nilai dari matriks kovarian, selanjutnya akan dilakukan perhitungan dekomposisi aigen (*eigen-decomposition*) untuk mendapatkan nilai *eigen value* (λ) dan *eigen vector* (v) (Yaicharoen and Yamada, 2021).

4. Dataset Partition

Langkah selanjutnya pada pemodelan *machine learning* prediksi kebangkrutan adalah *dataset partition*. Pada tahapan ini dilakukan pembagian dataset sebanyak 80% sebagai data *training* (pelatihan) dan sisanya 20% sebagai data *testing* (pengujian).

5. Tahapan Pelatihan (*Training phase*)

Tahapan ini merupakan fondasi dasar untuk membangun model machine learning prediksi kebangkrutan Perusahaan. Dataset yang terdiri dari kelas atribut bangkrut dan tidak bangkrut akan dipelajari polanya oleh sistem berdasarkan variabel yang digunakan (Tabel 1). Model mempelajari variabel yang ada guna menyesuaikan parameter dan mengurangi kesalahan prediksi.

6. Tahapan Pengujian *(Testing Phase)*

Proses testing berguna untuk mengukur kinerja akhir dari model prediksi kebangkrutan perusahaan. Tahapan ini dilakukan setelah proses pemodelan *machine learning* menggunakan data *training* selesai. Hasil dari proses *testing* akan memberikan evaluasi yang objektif berdasarkan parameter akurasi, presisi, *recall* ataupun kinerja secara keseluruhan.

7. Analisis Kinerja

Perhitungan kinerja masing-masing algoritma *machine learning* yang digunakan, dihitung dengan metode *confusion matrix*. *Confusion Matrix* adalah tabel yang digunakan untuk mengukur performa model klasifikasi dalam *machine learning*. Tabel ini membandingkan hasil prediksi model dengan nilai aktual (*ground truth*) dari dataset (Dennis et al., 2022).

Metode *confusion matrix* melihat perbandingan hasil dari prediksi model *machine learning* dengan model sebenarnya (Dasilas and Rigani, 2024). Pada metode ini, terdapat empat kombinasi nilai yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FP). **Tabel 4** menunjukan kombinasi dari nilai tersebut.

Tabel 4. *Confusion Matrix*

		Nilai Sebenarnya (Aktual)	
		Positive	***Negative***
Prediksi	***Positive***	*True Positive (TP)*	*False Positive (FP)*
	Negative	*False Negative (FN)*	*True Negative (TN)*

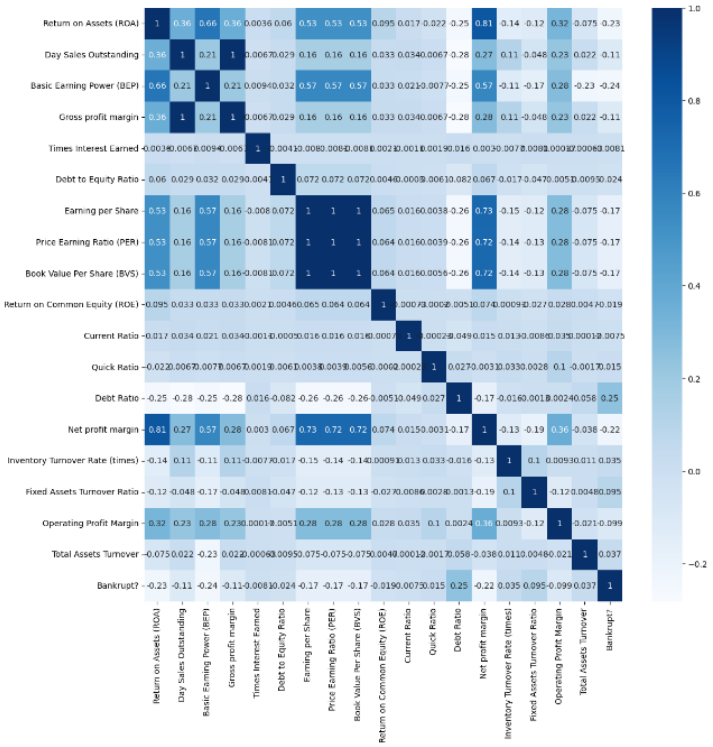
Nilai yang didapatkan dari *confusion matrix* akan digunakan untuk menentukan nilai dari indicator kinerja algoritma *machine learning* yaitu presisi, *recall*, akurasi dan dan *F1-Score*. Adapun masing-masing indicator dirumuskan sebagai berikut.

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (4)$$

$$F1-Score = 2\ x\ \frac{Presisi\ x\ Recall}{Presisi+Recall} \quad (5)$$



Gambar 2. Heatmap

4. HASIL DAN PEMBAHASAN

4.1 Evaluasi Dataset Penelitian

Proses prediksi kebangkrutan Perusahaan, hubungan antara rasio keuangan merupakan aspek penting yang perlu dianalisis. Hubungan antara variabel keuangan seperti profitabilitas, struktur modal, likuiditas dan risiko kebagkrutan menjadi aspek yang penting untuk menilai Kesehatan suatu Perusahaan. Heatmap pada **Gambar 2** menunjukan bagaimana keterkaitan variabel rasio keuangan mempengaruhi prediksi sistem menentukan kebangkrutan sebuah Perusahaan.

Berdasarkan hasil heatmap menunjukan adanya hubungan yang negatif antara rasio keuangan profitabilitas dengan risiko kebangkrutan. Dimana rasio Return on Assets (ROA) dan Net Profit Margin bernilai -0,25 dan -0,22 terhadap kebangkrutan. Nilai negatif ini menunjukan semakin tinggi keuntungan suatu perusahaan maka menurunkan rasio nilai kebangkrutan perusahaan tersebut. Rasio profitabilitas ini memberikan gambaran bahwa perusahaan yang efisien dalam mengelola keuntungan dan aset membuat perusahaan mampu mempertahankan keuangan yang lebih baik. Dengan kata lain, perusahaan yang kuat secara aset dan mampu mengelola laba secara finansial akan baik dan jauh dari prediksi bangkrut. Pada sisi lain, hubungan raiso antar Operating Profit Margin dan Net Profit Margin yang bernilai 0.72 menunjukan kemampuan perusahaan yang mampu mengenalikan biaya operasional berdampak pada laba bersih dan kestabilan keuangan perusahaan.

Pada variabel lain, struktur modal perusahaan yang tidak sehat memberikan dampak signifikan terhadap risiko kebangkrutan. Rasio Debt memiliki hubungan terhadap kebangkrutan sebesar 0.25. Hal ini menunjukan jika porsi utang yang dimiliki perusahan semakin meningkat, maka akan menyebabkan semakin besarnya potensi kebangkrutan. Karena perusahan akan mengalami kenaikan potensi gagal bayar. Rasio keuangan yang memiliki hubungan signifikan juga ditunjukan pada *Deb to Equity* terhadap *Debt Ratio* yaitu sebesar 0.82. Angka ini menunjukan utang yang besar dimiliki perusahaan akan memberikan beban finansial yang berlebih pada perusahaan.

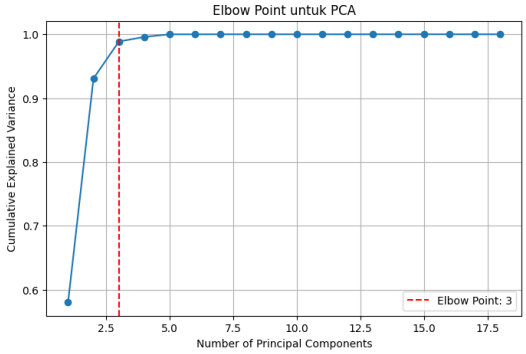
Perusahaan yang operasionalnya bergantung pada besaran hutang yang tidak diimbangi dengan profitabilitas akan menyebabkan meningkatnya potensi bangkrut. Dari nilai *heatmap* menunjukan tidak semua variabel berhubungan signifikan terhadap kebangktuan. hal ini bisa terlihat pada variable *Current Ratio* (0.075) dan *Quick Ratio* (0.015) menunjukkan korelasi yang sangat rendah terhadap kebangkrutan, yang mengindikasikan bahwa meskipun likuiditas penting untuk operasional jangka pendek, faktor ini bukan penentu utama dalam kelangsungan hidup perusahaan dalam jangka panjang. Sebaliknya, profitabilitas dan pengelolaan utang lebih berperan dalam menentukan apakah suatu perusahaan dapat bertahan atau mengalami kegagalan finansial.

Nilai rasio *earning per share* (EPS), *price earning ratio* (PER), dan *book value per share* (BVS) pada heatmap sebesar 1.00. Angka ini menunjukan hubungan yang signifikan antar ketiga

rasio tersebut dengan kebangkrutan perusahaan. Dimana, perusahan yang kinerja keuangannya baik akan berdampak pada nilai saham di pasar. Semakin baik atau tinggi nilai EPS dapat dikatakan perusahan semakin baik laba yang didapatkan perusahaan. Hal ini dapat menjadi daya tarik bagi perusahaan untuk mendapatkan investor. Dengan memahami hubungan antar variabel pada heatmap, peneliti bisa hubungan antar variabel dengan kelas kebangkrutan.

4.2 Metode PCA

Metode *Principal Component Analysis* (PCA) Dalam machine learning digunakan untuk mereduksi fitur data yang terlalu besar. Gambar 2 menunjukan hasil dari proses penerapan PCA pada dataset penelitian.



Gambar 3. PCA hasil reduksi dataset penelitian

Grafik yang ditunjukan pada **Gambar 3** mempresentasikan varian dataset asli dapat dijelaskan dengan proses reduksi dari sejumlah *Principal Components* (PCs). Jumlah dari PCs diwakili oleh sumbu X dan varian dataset diwakili oleh sumbu Y. Nilai *Elbow point* didapatkan pada saat komponen ke-3 yang ditandai garis merah putus-putus. PCs ke-3 dari proses PCA ini mampu menjelaskan 98,87% informasi yang terdapat dalam semua variabel penelitian. Guna melihat kontribusi variabel terhadap 3 PCs yang didapatkan peneliti menggunakan *Loading Matrix* (*Eigenvector*) yang ditunjukan pada Gambar 4.

Loading matrix atau *eigenvector* berguna untuk melihat sejauh mana variabel penelitian mempengaruhi ke-3 PCs yang dihasilkan. Interpretasi nilai yang dihasilkan pada loading matrix dilihat dari nilai, yaitu semakin tinggi nilai yang didapatkan (positif atau negatif) berarti variabel tersebut berkontribusi besar pada PCs. Dan semakin mendekati nol maka variabel tersebut tidak memiliki pengaruh terhadap PCs. berdasarkan hasil *Loading Matrix* (Gambar 4) komponen rasio keuangan yang berupa rasio profitabilitas dan likuiditas seperti *return on assets* (ROA), *earning per share* (EPS), *price earning ratio* (PER), *book value per share* (BVS), dan *return on common equity* (ROE) memiliki nilai PCs yang tinggi pada komponen PCs pertama (PC1). Ini menunjukan perusahaan yang rasio kinerja keuangannya baik cenderung berada pada PC1.

	PC1	PC2	PC3
Return on Assets (ROA)	0.346510	0.071830	-0.092061
Day Sales Outstanding	0.197485	0.587160	-0.136836
Basic Earning Power (BEP)	0.321632	-0.063611	-0.028672
Gross profit margin	0.198586	0.586308	-0.136951
Times Interest Earned	-0.002397	-0.009831	-0.037149
Debt to Equity Ratio	0.045188	-0.000869	0.108129
Earning per Share	0.394153	-0.215150	0.071397
Price Earning Ratio (PER)	0.394337	-0.214759	0.071787
Book Value Per Share (BVS)	0.394318	-0.214662	0.071396
Return on Common Equity (ROE)	0.040010	0.000642	-0.047914
Current Ratio	0.055798	0.137864	0.648728
Quick Ratio	0.002024	0.014902	-0.135729
Debt Ratio	-0.167076	-0.250045	-0.553178
Net profit margin	0.382756	-0.061917	-0.092946
Inventory Turnover Rate (times)	-0.065760	0.249028	0.039155
Fixed Assets Turnover Ratio	-0.089709	0.057232	0.117112
Operating Profit Margin	0.190480	0.085229	-0.388734
Total Assets Turnover	-0.043192	0.074512	-0.047201

Gambar 4. *Loading Matrix*

Pada PCs ke-2 (PC2) variabel yang berkontribusi terbesar terdapat pada variabel *day sales outstanding*, *inventory turnover rate*, *current ratio*, dan variabel *fixed assets turnover ratio*. Variabel pada PC2 lebih menggambarkan rasio keuangan yang berkaitan dengan efisiensi operasional dan perputaran aset perusahaan. Sedangkan pada PCs ke-3 (PC3) menggambarkan hutang tinggi pada sebuah perusahan berefek terhadap laba operasi yang semakin rendah. Hal ini bisa disebabkan karena besarnya beban bunga hutang yang harus dibayar perusahan. Pada PC3 variabel yang berkontribusi besar terdapat pada variabel *current ratio*, *debt ratio*, dan *operating profit margin*.

4.3 Performa Analisis Sistem Prediksi Kebangkrutan

Pada penelitian yang telah dilakukan, peneliti melakukan komparasi atau perbandingan performa beberapa algoritma klasifikasi untuk mendeteksi kebangkrutan perusahaan. Algoritma klasifikasi yang digunakan dalam penelitian ini antara lain algoritma KNN, SVM, Decision Tree dan Naive Bayes. Setiap algoritma dilakukan pengujian performa sebelum menggunakan metode *oversampling* dan pengujian setelah penerapan metode *oversampling*. Perbandingan ini dilakukan untuk melihat sejauh mana dampak pengaruh metode *oversampling* terhadap performa algoritma keseluruhan.

Dari Tabel 5, setelah penerapan metode *oversampling* performa algoritma mengalami perubahan yang signifikan, terutama pada peningkatan nilai *recall* untuk kelas minoritas. Pada sisi lain, peningkatan nilai *recall* ini berdampak pada penurunan nilai precision dan akurasi sistem prediksi secara keseluruhan.

Tabel 1. Performa Algoritma

Metode	Kelas	Tidak menggunkan *Random Oversampling*			Menggunakan *Random Oversampling*		
		Precision	*Recall*	F1-Score	Precision	*Recall*	F1-Score
KNN	0	0.96	1.00	0.98	0.66	0.94	0.78
	1	0.50	0.02	0.04	0.90	0.51	0.65
	Akurasi		**0.96**			**0.73**	
	Macro Avg	0.73	0.51	0.51	0.78	0.73	0.71
	Weighted Avg	0.95	0.96	0.95	0.78	0.73	0.71
Naive Bayes	0	0.97	0.99	0.98	0.73	0.32	0.45
	1	0.31	0.16	0.21	0.57	0.88	0.69
	Akurasi		**0.96**			**0.65**	
	Macro Avg	0.64	0.57	0.60	0.65	0.60	0.57
	Weighted Avg	0.95	0.96	0.95	0.65	0.60	0.57
SVM	0	0.96	1.00	0.98	0.78	0.87	0.83
	1	0.00	0.00	0.00	0.86	0.76	0.80
	Akurasi		**0.96**			**0.82**	
	Macro Avg	0.48	0.50	0.49	0.82	0.82	0.81
	Weighted Avg	0.95	0.96	0.95	0.82	0.82	0.81
Decision Tree	0	0.97	0.92	0.94	0.51	1.00	0.67
	1	0.89	0.27	0.49	1.00	0.27	0.38
	Akurasi		**0.89**			**0.62**	
	Macro Avg	0.54	0.59	0.55	0.71	0.52	0.38
	Weighted Avg	0.94	0.89	0.91	0.71	0.52	0.38

Keterangan : 0 = Kelas Tidak Bangkrut; 1 = Kelas Bangkrut.

Nilai akurasi algoritma setelah penerapan metode imbalanced dataset mengalami penurunan, seperti pada algoritma SVM sebelum penerapan *oversampling* mendapatkan akurasi sebesar 0.96 dan mengalami penurunan menjadi 0.82 setelah dilakukan penerapan metode *oversampling*. Penurunan akurasi ini dikarenakan pada penerapan metode *oversampling* memaksa algoritma untuk bekerja lebih optimal dalam mempelajari perbedaan kelas dataset, dan tidak tergantung hanya pada kelas mayoritas saja.

Hasil penelitian menunjukkan bahwa kombinasi metode *Random Oversampling* dan *Principal Component Analysis* (PCA) secara signifikan meningkatkan kemampuan model *machine learning* dalam memprediksi kebangkrutan perusahaan, terutama dalam mengatasi ketidakseimbangan kelas (*imbalanced dataset*). Sebelum penerapan *oversampling*, algoritma cenderung bias terhadap kelas mayoritas (perusahaan tidak bangkrut), seperti ditunjukkan oleh nilai *recall* yang sangat rendah pada kelas minoritas (kebangkrutan). Misalnya, algoritma *Support Vector Machine* (SVM) awalnya menghasilkan *recall* 0,00 untuk kelas bangkrut, sementara *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) masing-masing hanya mencapai 0,16 dan 0,02. Setelah diterapkan *Random Oversampling*, terjadi peningkatan drastis pada *recall* kelas minoritas: SVM meningkat menjadi 0,76, KNN menjadi 0,51, dan Naïve Bayes mencapai 0,88. Namun, peningkatan ini diiringi penurunan akurasi keseluruhan (misalnya, akurasi SVM turun dari 0,96 menjadi 0,82), yang mengindikasikan model tidak lagi didominasi oleh kelas mayoritas.

PCA berhasil mereduksi dimensi dataset hingga 98,87% varian data tetap terjaga, dengan menggabungkan variabel keuangan seperti *Return on Assets* (ROA), Debt Ratio, dan *Operating Profit Margin* ke dalam tiga komponen utama (*Principal Components*). Reduksi ini mengurangi kompleksitas komputasi dan mengatasi masalah *multicollinearity* antar variabel. Sementara itu, *Decision Tree* gagal memanfaatkan oversampling secara optimal, dengan recall kelas minoritas tetap stagnan di 0,27 dan penurunan F1-score dari 0,49 menjadi 0,38, menandakan gejala *overfitting*.

Secara keseluruhan, SVM menjadi algoritma terbaik dengan keseimbangan *precision* (0,86), *recall* (0,76), dan *F1-score* (0,80), sementara *Decision Tree* tidak direkomendasikan untuk dataset tidak seimbang. Hasil ini menegaskan bahwa kombinasi PCA dan Random Oversampling efektif meningkatkan sensitivitas model terhadap kelas minoritas, meskipun mengorbankan akurasi absolut, yang krusial untuk deteksi dini kebangkrutan dalam konteks Indonesia.

Perubahan performa masing-masing algoritma dalam memprediksi kebangkrutan setelah dilakukan penerapan *oversampling* dijelaskan berikut ini.

Pengaruh Random oversampling terhadap Algoritma KNN

Tabel 2. Pengaruh *Random Oversampling* terhadap KNN

Metrik	Sebelum *Oversampling*	Sesudah *Oversampling*	Perubahan
Accuracy	0.96	0.73	Menurun
Precision (kelas 1)	0.50	0.90	Meningkat signifikan
Recall (kelas 1)	0.02	0.51	Meningkat signifikan
F1-Score (kelas 1)	0.04	0.65	Meningkat

Tabel 6 menunjukan pengaruh penerapan metode *random oversampling* terhadap Algoritma KNN sebelum diterapkan *Oversampling* menghasilkan performa yang sangat bias terhadap kelas minoritas (kelas 1 = kelas bangkrut). Pada kelas minoritas menghasilkan *recall* sebesar 0.02, hal ini menunjukan akurasi algoritma yang sebesar 0.96 hanya mampu mengklasifikasikan kelas

mayoritas dibandingkan kelas minoritas. Setelah diterapkan metode *oversampling*, akurasi algoritma mengalami penurunan menjadi 0.73 tapi nilai *recall* meningkat menjadi 0.51. Pada algoritma KNN untuk kasus prediksi kebangkrutan perusahaan metode *Oversampling* mampu meningkatkan kemampuan algoritma dalam mengenali kelas minoritas. Peningkatan nilai *recall* menyebabkan penurunan nilai akurasi dikarenakan adanya distribusi kelas yang lebih seimbang.

Pengaruh Random Oversampling terhadap Algoritma Naive Bayes

Tabel 3. Pengaruh *Random Oversampling* terhadap Naïve Bayes

Metrik	Sebelum *Oversampling*	Sesudah *Oversampling*	Perubahan
Accuracy	0.96	0.65	Menurun
Precision (kelas 1)	0.31	0.57	Meningkat
Recall (kelas 1)	0.16	0.88	Meningkat signifikan
F1-Score (kelas 1)	0.21	0.69	Meningkat

Tabel 7 menunjukan performa algoritma Naïve Bayes sebelum dan sesudah penerapan metode *Random Oversampling*. Pada algoritma Naive Bayes, sebelum diterapkan metode *Oversampling* menghasilkan akurasi sebesar 0.96 tapi nilai *recall* untuk kelas 1 (kelas bangkrut) hanya sebesar 0.16. Hal ini berarti model klasifikasi hanya mampu mengenali dataset kelas mayoritas (tidak bangkrut). Setelah dilakukan penerapan metode *oversampling*, akurasi model menurun menjadi 0.65 dan *recall* mengalami peningkatan yang signifikan menjadi 0.88. *Oversampling* pada algoritma *naive bayes* untuk klasifikasi kebangkrutan perusahaan menurunkan nilai akurasi. Namun, metode ini mampu meningkatkan nilai *recall* sehingga keseimbangan antar kelas dalam melakukan prediksi menjadi meningkat.

Pengaruh Random Oversampling terhadap Algoritma SVM

Tabel 4. Pengaruh *Random Oversampling* terhadap SVM

Metrik	Sebelum *Oversampling*	Sesudah *Oversampling*	Perubahan
Accuracy	0.96	0.82	Menurun
Precision (kelas 1)	0.00	0.86	Meningkat signifikan
Recall (kelas 1)	0.00	0.76	Meningkat signifikan
F1-Score (kelas 1)	0.00	0.80	Meningkat signifikan

Tabel 8 menunjukan performa algoritma SVM sebelum dan sesudah penerapan metode *Random Oversampling*. Algoritma SVM memiliki nilai akurasi 0.92 sebelum diterapkan oversampling. Namun, hasil dari akurasi ini tidak mampu mengenali kelas minoritas yang bisa dilihat dari nilai precision, *recall* dan F1-Score pada kelas 1 bernilai 0.00. setelah penerapan metode oversampling, nilai *recall* pada kelas 1 (kelas bangkrut) meningkat menjadi 0.86. Akurasi model mengalami penurunan, seiring dengan seimbangnya disktribusi kelas.

Pengaruh Random Oversampling terhadap Decision Tree

Tabel 5. Pengaruh *Oversampling* terhadap *Decision Tree*

Metrik	Sebelum *Oversampling*	Sesudah *Oversampling*	Perubahan
Accuracy	**0.89**	**0.62**	Menurun
Precision (kelas 1)	**0.89**	**1.00**	Meningkat
Recall (kelas 1)	**0.27**	**0.27**	Tetap sama
F1-Score (kelas 1)	**0.49**	**0.38**	Menurun

Tabel 9 menunjukan nilai performa model klasifikasi dengan algoritma *decision tree* sebelum dan sesudah penerapa metode *randm oversampling*. Dari hasil pengujian, algoritma decision tree tidak mampu mengoptimalkan metode oversampling dalam meningkatkan kinerja algoritma. Hal ini dapat terlihat dari penurunan akurasi, walaupun nilai precision mengalami peningkatan. Nilai *recall* setelah dan sebelum penerapan metode sama yaitu 0.27, hal ini membuktikan algoritma tetap tidak mampu mengoptimalkan memprediksi dataset minoritas.

5. KESIMPULAN

Dari hasil penelitian menunjukan metode *oversampling* memberikan dampak yang signifikan untuk meningkatkan performa algoritma klasifikasi pada kasus prediksi kebangkrutan perusahaan. Metode ini mampu meningkatkan nilai *recall* pada kelas minoritas, yang sebelum penerapan kelas mayoritas lebih dominan. Algoritma *K-Nearest Neighbors* (KNN), *naïve bayes*, dan SVM menghasilkan peningkatan nilai *recall* setelah menerapkan metode *oversampling*. Pada algoritma *decision tree* setelah penerapan *oversampling* mengalmi permasalahan *overfitting*. Hal ini dikarenakan pada algoritma decision tree mengeneralisasi dataset yang telah di *oversample*, sehingga algoritma menghasilkan lebih banyak false positif. Algoritma yang paling stabil untuk penerapan *oversampling* adalah algoritma SVM, dimana nilai *precision* dan *recall* seimbang, serta nilai *f1-score* yang tinggi.

DAFTAR PUSTAKA

ADEBUSOLA, S.O., OWOLAWI, P.A., OJO, J.S., MASWIKANENG, P.S. AND AYO, A.O., 2025. Application of principal component analysis and Artificial neural networks for the prediction of QoS in FSO links over South Africa. *Results in Optics*, 19, p.100796. https://doi.org/10.1016/J.RIO.2025.100796.

ALJAWAZNEH, H., MORA, A., GARCÍA-SÁNCHEZ, P. AND CASTILLO, P., 2021a. Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure. *IEEE Access*, PP, p.1. https://doi.org/10.1109/ACCESS.2021.3093461.

ALJAWAZNEH, H., MORA, A.M., GARCÍA-SÁNCHEZ, P. AND CASTILLO-

VALDIVIESO, P.A., 2021b. Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure. *IEEE Access*, 9, pp.97010–97038. https://doi.org/10.1109/ACCESS.2021.3093461.

BINANTO, I., JAMLU, M.S., WIBISONO, R.D.A. AND SIANIPAR, N.F., 2024. A Comparison of Random Forest and Support Vector Machine Classification Algorithms for Imbalanced and Balanced Rodent Tuber Dataset with Random Oversampling Method. In: *2024 Ninth International Conference on Informatics and Computing (ICIC)*. pp.1–4. https://doi.org/10.1109/ICIC64337.2024.10956313.

BYUN, J., LEE, J., LEE, H. AND SON, B., 2025. Balancing Explainability and Privacy in Bank Failure Prediction: A Differentially Private Glass-Box Approach. *IEEE Access*, 13, pp.1546–1565. https://doi.org/10.1109/ACCESS.2024.3523967.

CHEN, L., 2023. Machine Learning-based Analysis and Prediction of Telecoms Customer Churn. In: *2023 5th International Conference on Applied Machine Learning (ICAML)*. pp.122–126. https://doi.org/10.1109/ICAML60083.2023.00032.

DAS, R., BISWAS, S.K., DEVI, D. AND SARMA, B., 2020. An Oversampling Technique by Integrating Reverse Nearest Neighbor in SMOTE: Reverse-SMOTE. In: *2020 International Conference on Smart Electronics and Communication (ICOSEC)*. pp.1239–1244. https://doi.org/10.1109/ICOSEC49089.2020.9215387.

DASILAS, A. AND RIGANI, A., 2024. Machine learning techniques in bankruptcy prediction: A systematic literature review. *Expert Systems with Applications*, 255, p.124761. https://doi.org/10.1016/J.ESWA.2024.124761.

DENNIS, BUDIANTO, I.R., AZARIA, R.K. AND GUNAWAN, A.A.S., 2022. Machine Learning-based Approach on Dealing with Binary Classification Problem in Imbalanced Financial Data. In: *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*. pp.152–156. https://doi.org/10.1109/ISMODE53584.2022.9742834.

DIANA, H. AND HIDAYAT, D., 2023. Deteksi Kecurangan Laporan Keuangan dan Prediksi Kebangkrutan Perusahaan Sebelum dan Saat Pandemi Covid-19 dengan Menggunakan Perbandingan Pengukuran Model Altman Z – Score, Grover, Springate dan Zmijewski. *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan*, [online] 17(1), pp.255–278. https://doi.org/10.35931/AQ.V17I1.1801.

GALIL, K., HAUPTMAN, A. AND ROSENBOIM, R.L., 2023. Prediction of corporate credit ratings with machine learning: Simple interpretative models. *Finance Research Letters*, 58, p.104648. https://doi.org/10.1016/J.FRL.2023.104648

GURNANI, I., VINCENT, TANDIAN, F. AND ANGGREAINY, M., 2021. *Predicting Company Bankruptcy Using Random Forest Method*. https://doi.org/10.1109/AiDAS53897.2021.9574384.

HE, F. AND ZHANG, Z., 2020. Nonlinear Fault Detection of Batch Processes Using Functional Local Kernel Principal Component Analysis. *IEEE Access*, 8, pp.117513–117527. https://doi.org/10.1109/ACCESS.2020.3004564.

IPARRAGUIRRE-VILLANUEVA, O. AND CABANILLAS-CARBONELL, M., 2024. Predicting business bankruptcy: A comparative analysis with machine learning models. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(3), p.100375. https://doi.org/10.1016/J.JOITMC.2024.100375.

JASIM, Z.A., ZAHID, Z., UL-SAUFIE, A.Z. AND MANSOR, M.M., 2024. Comparison between Principal Component Analysis and Sparse Principal Component Analysis as Dimensional Reduction Techniques for Random Forest based High Dimensional Data Classification. In: *2024 IEEE International Conference on Computing (ICOCO)*. pp.7–11. https://doi.org/10.1109/ICOCO62848.2024.10928248.

KAIB, M.T.H., KOUADRI, A., HARKAT, M.-F., BENSMAIL, A. AND MANSOURI, M., 2024. Improvement of Kernel Principal Component Analysis-Based Approach for Nonlinear Process Monitoring by Data Set Size Reduction Using Class Interval. *IEEE Access*, 12, pp.11470–11480. https://doi.org/10.1109/ACCESS.2024.3354926.

LIU, C., JIN, S., WANG, D., LUO, Z., YU, J., ZHOU, B. AND YANG, C., 2022. Constrained Oversampling: An Oversampling Approach to Reduce Noise Generation in Imbalanced Datasets With Class Overlapping. *IEEE Access*, 10, pp.91452–91465. https://doi.org/10.1109/ACCESS.2020.3018911.

LORD, J., LANDRY, A., SAVAGE, G.T. AND WEECH-MALDONADO, R., 2020. Predicting Nursing Home Financial Distress Using the Altman Z-Score. *Inquiry (United States)*, 57. https://doi.org/10.1177/0046958020934946

MASDIANTINI, P.R. AND WARASNIASIH, N.M.S., 2020. Laporan Keuangan dan Prediksi Kebangkrutan Perusahaan. *JIA (Jurnal Ilmiah Akuntansi)*, [online] 5(1), pp.196–220. https://doi.org/10.23887/JIA.V5I1.25119.

PAMUNGKAS, S., 2023. Financial Distress Analysis Using the Ohlson Model in Indonesian State Owned Enterprises. *Journal of Accounting and Finance Management*, [online] 3(6), pp.272–381. https://doi.org/10.38035/JAFM.V3I6.176.

PAPÍK, M. AND PAPÍKOVÁ, L., 2024. Automated Machine Learning in Bankruptcy Prediction of Manufacturing Companies. *Procedia Computer Science*, 232, pp.1428–1436. https://doi.org/10.1016/J.PROCS.2024.01.141.

PARK, M.S., SON, H., HYUN, C. AND HWANG, H.J., 2021a. Explainability of Machine Learning Models for Bankruptcy Prediction. *IEEE Access*, 9, pp.124887–124899. https://doi.org/10.1109/ACCESS.2021.3110270.

PARK, S., SON, H., HYUN, C. AND HWANG, H., 2021b. Explainability of Machine Learning Models for Bankruptcy Prediction. *IEEE Access*,

PP, p.1. https://doi.org/10.1109/ACCESS.2021.3110270.

PRANAVI, N.S.S., SRUTHI, T.K.S.S., SIRISHA, B.J.N., Nayak, M.S. and Thadikemalla, V.S.G., 2022. Credit Card Fraud Detection Using Minority Oversampling and Random Forest Technique. In: *2022 3rd International Conference for Emerging Technology (INCET)*. pp.1–6. https://doi.org/10.1109/INCET54531.2022.9824146.

PUTRI, M.E. AND CHALLEN, A.E., 2021. Prediksi Kebangkrutan Pada Perusahaan Yang Terdaftar Di Bursa Efek Indonesia. *JAS (Jurnal Akuntansi Syariah)*, [online] 5(2), pp.126–141. https://doi.org/10.46367/JAS.V5I2.425.

RAHAYU, N.E.E. AND USMANSYAH, U., 2021a. Altman Z-Score Model to Analyze Bankruptcy of Islamic Commercial Bank POJK No. 12/POJK.03/2020. *LAA MAISYIR : Jurnal Ekonomi Islam*, [online] 8(2), pp.173–191. https://doi.org/10.24252/LAMAISYIR.V8I2.23097.

RAHAYU, N.E.E. AND USMANSYAH, U., 2021b. Altman Z-Score Model to Analyze Bankruptcy of Islamic Commercial Bank POJK No. 12/POJK.03/2020. *LAA MAISYIR : Jurnal Ekonomi Islam*, [online] 8(2), pp.173–191. https://doi.org/10.24252/LAMAISYIR.V8I2.23097.

SETIAWAN, F., 2021. Financial Distress Analysis Using Altman Z-Score Model In Sharia Banking In Indonesia. *IQTISHODUNA: Jurnal Ekonomi Islam*, [online] 10(2). https://doi.org/10.36835/IQTISHODUNA.V10I2.938.

SETO, A.A., 2022. Altman Z-Score Model, Springate, Grover, Ohlson and Zmijweski to Assess the Financial Distress Potential of PT. Garuda Indonesia Tbk During and After the Covid-19 Pandemic. *Enrichment : Journal of Management*, [online] 12(5), pp.3819–3826. https://doi.org/10.35335/ENRICHMENT.V12I5.923.

SHETTY, S., MUSA, M. AND BRÉDART, X., 2022. Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management 2022, Vol. 15, Page 35*, [online] 15(1), p.35. https://doi.org/10.3390/JRFM15010035.

TARAWNEH, A.S., HASSANAT, A.B., ALTARAWNEH, G.A. and Almuhaimeed, A., 2022. Stop Oversampling for Class Imbalance Learning: A Review. *IEEE Access*, 10, pp.47643–47660. https://doi.org/10.1109/ACCESS.2022.3169512.

WIDIASMARA, A. AND RAHAYU, H.C., 2019. PERBEDAAN MODEL OHLSON, MODEL TAFFLER DAN MODEL SPRINGATE DALAM MEMPREDIKSI FINANCIAL DISTRESS. *INVENTORY: JURNAL AKUNTANSI*, [online] 3(2), pp.141–158. https://doi.org/10.25273/INVENTORY.V3I2.5242.

XIE, S., LIN, H., MA, T., PENG, K. AND SUN, Z., 2024. Prediction of joint roughness coefficient via hybrid machine learning model combined with principal components analysis. *Journal of Rock Mechanics and Geotechnical Engineering*. https://doi.org/10.1016/J.JRMGE.2024.05.059.

XUE, K., QI, Y., DUAN, H., CAO, A. AND WANG, A., 2024. Prediction of coal and gas outburst hazard using kernel principal component analysis and an enhanced extreme learning machine approach. *Geohazard Mechanics*, 2(4), pp.279–288. https://doi.org/10.1016/J.GHM.2024.09.002.

YAICHAROEN, A. AND YAMADA, K., 2021. Improving Support Vector Classification Efficiency with Principal Component Analysis. In: *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. pp.862–865. https://doi.org/10.1109/ECTI-CON51831.2021.9454883.