

# Penerapan Regresi Logistik dan Analisis Diskriminan dalam Klasifikasi Status Kebangkrutan Korporat

Widya Louisa<sup>1</sup>, Melinda Venta Lydia Siburian<sup>2</sup>, Khalila Salma Dzakia Agram<sup>3</sup>, Ulfa Siti Nuraini, S.Stat., M.Stat.<sup>4</sup>

<sup>1</sup>Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: [widya.23180@mhs.unesa.ac.id](mailto:widya.23180@mhs.unesa.ac.id)

<sup>2</sup>Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: [Melinda.23124@mhs.unesa.ac.id](mailto:Melinda.23124@mhs.unesa.ac.id)

<sup>3</sup>Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: [khalila.23126@mhs.unesa.ac.id](mailto:khalila.23126@mhs.unesa.ac.id)

<sup>4</sup>Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: [ulfanuraini@unesa.ac.id](mailto:ulfanuraini@unesa.ac.id)

**ABSTRAK** - Kemampuan untuk memprediksi kebangkrutan perusahaan menjadi aspek krusial dalam pengambilan keputusan finansial, terutama di tengah dinamika pasar dan ketidakpastian ekonomi. Penelitian ini bertujuan untuk mengklasifikasikan status kebangkrutan perusahaan di Italia dengan membandingkan dua metode statistik klasik yang umum digunakan, yaitu *Regresi Logistik* (LR) dan *Linear Discriminant Analysis* (LDA). Dataset yang digunakan mencakup 96 indikator keuangan dari 6.819 perusahaan, dengan status kebangkrutan sebagai variabel target biner. Untuk mengatasi multikolinearitas dan kompleksitas data berdimensi tinggi, dilakukan transformasi menggunakan *Principal Component Analysis* (PCA), dan selanjutnya hanya 10 komponen utama digunakan dalam proses klasifikasi. Model dibangun setelah uji asumsi dan pembersihan *outlier* dilakukan. Hasil evaluasi model menggunakan *confusion matrix*, akurasi, *sensitivitas*, spesifisitas, serta kurva ROC menunjukkan bahwa Regresi Logistik menghasilkan akurasi tinggi (97,5%) namun memiliki sensitivitas rendah terhadap kelas minoritas. Sementara itu, LDA menunjukkan performa klasifikasi yang baik secara keseluruhan (akurasi >96%), namun kesulitan dalam membedakan perusahaan bangkrut akibat dominasi kelas mayoritas. Temuan ini menekankan pentingnya penyeimbangan data dan pemilihan metode klasifikasi yang tepat. Regresi Logistik terbukti lebih fleksibel karena tidak membutuhkan asumsi distribusi, namun LDA tetap memberikan interpretasi klasifikasi yang baik jika asumsi terpenuhi. Kedua metode dapat berkontribusi dalam pengembangan sistem deteksi dini risiko kebangkrutan yang akurat dan praktis di sektor korporat.

**Keywords** — *Kebangkrutan, Linear Discriminant analysis, Logistic Regression*

## 1. PENDAHULUAN

Analisis multivariat merupakan metode statistik yang digunakan untuk menganalisis data yang memiliki banyak variabel yang saling berkorelasi. Salah satu cabang utama dari analisis multivariat adalah analisis

dependensi, yaitu analisis yang melibatkan satu variabel respon dan beberapa variabel prediktor. Dua metode yang sering digunakan dalam analisis dependensi adalah *regresi logistik* dan *analisis diskriminan*.

Analisis diskriminan merupakan metode klasifikasi yang digunakan untuk memodelkan hubungan antara satu variabel respon kategorik (seperti status kebangkrutan: bangkrut atau tidak) dengan sejumlah variabel bebas kuantitatif. Metode ini mengasumsikan bahwa data pada masing-masing kelompok mengikuti distribusi normal multivariat dan memiliki struktur kovarians yang sama di setiap kelompok.

Fungsi diskriminan yang terbentuk digunakan untuk memisahkan dua atau lebih kelompok secara linier, dengan tujuan meminimalkan kesalahan klasifikasi.

Sementara itu, regresi logistik adalah metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel biner dengan satu atau lebih variabel bebas, baik kuantitatif maupun kategorik. Menurut Agresti (1990),

regresi logistik menghasilkan probabilitas kejadian suatu peristiwa (misalnya kebangkrutan) melalui transformasi logistik dari kombinasi linier variabel bebas. Metode ini tidak mengharuskan asumsi normalitas maupun kesamaan kovarians antar kelompok, sehingga sering dianggap lebih fleksibel daripada analisis diskriminan, khususnya dalam kondisi data yang tidak ideal.

Salah satu penerapan kedua metode ini adalah dalam studi klasifikasi status kebangkrutan perusahaan. Di tengah dinamika ekonomi global, prediksi kebangkrutan menjadi hal yang penting untuk menilai keberlanjutan usaha suatu perusahaan. Ketidakmampuan dalam mengelola likuiditas, menekan beban utang, atau menjaga profitabilitas dapat menyebabkan perusahaan mengalami krisis keuangan yang serius hingga berujung pada kebangkrutan. Dalam hal ini, pengembangan model prediktif untuk mengidentifikasi perusahaan yang berisiko bangkrut menjadi alat strategis dalam manajemen risiko.

Penelitian ini memanfaatkan data keuangan perusahaan Italia yang terdiri dari 96 indikator untuk mengklasifikasikan status kebangkrutan. Untuk mengatasi kompleksitas dan multikolinearitas antar variabel, digunakan teknik reduksi dimensi *Principal Component Analysis (PCA)*. Selanjutnya, dilakukan klasifikasi menggunakan dua pendekatan yaitu Regresi Logistik, dan Linear Discriminant Analysis (LDA), serta

evaluasi performa masing-masing model untuk menentukan metode mana yang memberikan ketepatan klasifikasi tertinggi.

## 2. KAJIAN PUSTAKA

### 2.1 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) adalah salah satu teknik reduksi dimensi yang dapat digunakan untuk menyederhanakan dataset tanpa kehilangan informasi yang signifikan. PCA bekerja dengan mentransformasikan variabel yang berpotensi berkorelasi menjadi sekumpulan variabel yang lebih kecil yang disebut komponen utama. PCA dikembangkan pertama kali oleh Karl Pearson pada tahun 1901 dan semakin berkembangnya teknologi PCA makin populer digunakan karena PCA sangat efektif dalam memvisualisasikan dan mengeksplorasi kumpulan data berdimensi tinggi, atau data dengan banyak fitur, karena dapat dengan mudah mengidentifikasi tren, pola, atau outlier

Langkah pertama dalam PCA adalah standarisasi, yaitu mengubah data agar memiliki rata-rata 0 dan variansi 1. Proses ini menggunakan rumus sebagai berikut:

$$z = \frac{x - \mu}{\sigma}$$

Keterangan :

$x$  = Nilai asli,

$\sigma$  = Standar deviasi fitur,

$\mu$  = Niali rata-rata fitur,

Langkah kedua yaitu menghitung kovarian antar fitur untuk memahami hubungan linear antar variabel.

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

Langkah ketiga menghitung eigenvalue dan eigenvector.

langkah keempat memilih komponen utama berdasarkan nilai eigenvalue tertinggi. biasanya, mengambil beberapa komponen pertama yang menjelaskan >80% dari total variansi.

langkah kelima memproyeksikan ke dimensi baru dengan cara mengalikan dengan eigenvector terpilih, hasil akhirnya merupakan dataset baru berdimensi lebih rendah namun tetap menyimpan informasi penting.

## 2.2 REGRESI LOGISTIK

Regresi logistik adalah metode statistik yang digunakan untuk memprediksi probabilitas kejadian kategorik, terutama dua kategori seperti “bangkrut” atau “tidak bangkrut”. Model ini memanfaatkan fungsi logit untuk menghubungkan variabel independen dengan log odds dari outcome:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Hasilnya sering ditafsirkan dalam bentuk odds ratio, yang menunjukkan pengaruh masing-masing variabel terhadap kemungkinan kebangkrutan.

### 2.2.1 Uji Asumsi Multikolinearitas

Multikolinearitas adalah masalah yang sering ditemukan pada model regresi, maka perlu dilakukan pendekatan lain agar tidak menghasilkan interpretasi koefisien regresi yang tidak tepat dan mungkin akan terjadi kesalahan saat pengambilan keputusan. Karena pada umumnya akan dilakukan tindakan membuang peubah prediktor yang saling berkorelasi cukup tinggi. Padahal kenyataannya peubah prediktor tersebut cukup berpengaruh terhadap peubah respon.

Multikolinearitas digunakan untuk menunjukkan adanya derajat linieritas yang tinggi di antara peubah-peubah prediktor. Jika multikolinieritas bersifat tinggi atau sempurna, maka koefisien regresi dari peubah prediktor tidak dapat ditentukan dan salah baku-nya tidak terhitung, yang dapat diartikan bahwa koefisien regresi tidak dapat diduga dengan tingkat keakuratan yang tinggi.

Multikolinieritas sering diduga jika nilai  $R^2$  dan korelasi yang tinggi antara peubah prediktor, tetapi sedikit koefisien regresi yang signifikan secara individu. Jadi secara parsial tidak mempunyai pengaruh terhadap peubah prediktor.

Cara untuk mengidentifikasi adanya multikolinieritas adalah dengan VIF (Variation Inflation Factor).

$$VIF_j = \frac{1}{1 - R_j^2}$$

dimana:

$R_j^2$  = Koefisien determinasi antara  $X_j$  dengan peubah prediktor lainnya ;  $j=1,2,\dots,p$

Apabila nilai koefisien determinasi mendekati 1, maka nilai VIF akan bertambah besar. Hal ini terjadi jika peubah prediktor yang lain. Peubah prediktor dikatakan terdapat multikolinieritas jika nilai VIF > 10.

Hamilton (1992) mengatakan bahwa pada data yang terjadi multikolinieritas akan dihasilkan pendugaan parameter yang kurang baik, galat dan ragam galat akan besar. Apabila ragam galat besar maka akan memperkecil statistik uji-t dan memperlebar selang kepercayaan bagi  $\beta_j$ .

Pendugaan parameter dengan metode maksimum likelihood pada model regresi logistik tidak dapat dilakukan karena matriks ragam peragamnya  $[X'VX]^{-1}$  diasumsikan non singular dan apabila di antara peubah prediktor terdapat korelasi maka  $|X'VX|=0$ .

Galat penduga koefisien regresi ( $\hat{\beta}$ ) nilainya akan semakin besar jika terjadi multikolinieritas. Hal ini dapat dilihat dari matriks ragam peragam  $[X'VX]^{-1}$  yang dibutuhkan untuk menentukan galat dari penduga koefisien regresi logistik melalui persamaan:

$$Se(\hat{\beta}) = \sqrt{Var(\hat{\beta}_j)}$$

di mana:

$Var(\hat{\beta}_j)$  = Elemen ke-j dari diagonal utama matriks  $[X'VX]^{-1}$

Masalah multikolinieritas pada regresi logistik akan terjadi apabila matriks  $|X'VX|$  bersifat singular atau mendekati singular ( $|X'VX| \rightarrow 0$ ), sehingga unsur-unsur matriks  $[X'VX]^{-1}$  yang diperoleh dari persamaan :

$$|X'VX|^{-1} = \frac{1}{|X'VX|} adj(X'VX)$$

akan semakin besar atau mendekati tak hingga nilai  $se(\hat{\beta}_j)$  juga akan semakin besar, maka multikolinieritas mengakibatkan presisi penduga koefisien regresi semakin kecil. Dengan demikian semakin besar nilai

$se(\hat{\beta}_j)$  maka nilai statistik W pada uji Wald semakin kecil. Dapat dilihat dari rumus W hitung untuk menguji koefisien regresi logistik secara parsial pada  $X_j$  sebagai berikut:

$$W_{hit} = \frac{|\hat{\beta}|}{se(\hat{\beta}_j)}$$

Nilai  $Se()$  akan semakin besar jika paling sedikit satu peubah prediktor berkorelasi terhadap peubah prediktor yang lainnya. Sehingga peluang  $P(x^2 > W^2) < \alpha$  akan semakin besar. Jadii adanya multikolinieritas akan menurunkan kekuatan uji Wald.

### 2.2.2 R-Square

Dalam regresi logistik, nilai  $R^2$  (pseudo- $R^2$ ) berfungsi sebagai ukuran sejauh mana model dapat menjelaskan variasi dalam data, meskipun tidak memiliki interpretasi yang sama persis seperti pada regresi linear. Beberapa jenis pseudo- $R^2$  yang umum digunakan adalah McFadden's  $R^2$ , Cox and Snell  $R^2$ , dan Nagelkerke  $R^2$ . Nilai pseudo- $R^2$  berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan kemampuan prediksi model yang lebih baik. Misalnya, McFadden (1974) menyarankan bahwa nilai sekitar 0.2 hingga 0.4 sudah dianggap baik dalam konteks model logistik. Dalam

praktik modern, sebagaimana dijelaskan Sarah Lee (2025), pseudo- $R^2$  sebaiknya digunakan bersamaan dengan metrik lain—seperti AUC, confusion matrix, atau Brier score—untuk memberikan gambaran yang lebih komprehensif terhadap kualitas dan validitas model regresi logistik

### 2.2.3 Odds ratio

Odds ratio (OR) adalah ukuran penting yang mengkuantifikasi seberapa besar peluang terjadinya sebuah peristiwa (misalnya kebangkrutan) seiring perubahan satu unit pada prediktor, dengan variabel lain dikontrol konstan. OR dihitung dengan mengubah koefisien logit ( $\beta$ ) menjadi  $e^{\beta}$ , sehingga  $OR > 1$  menunjukkan bahwa peningkatan variabel prediktor berasosiasi dengan peningkatan odds dari peristiwa tersebut, sementara  $OR < 1$  menunjukkan penurunan odds. Salah satu keunggulan OR adalah kestabilan efek lintas level X, sehingga efek dari variabel tetap persisten terukur (UVOR) (Norton, 2018).

### 2.2.4 Uji Signifikansi Model Regresi Logistik

Dalam regresi logistik, uji signifikansi model secara keseluruhan bertujuan untuk menilai apakah model yang dibangun memberikan

peningkatan yang signifikan dibandingkan model tanpa prediktor (model null). Salah satu pendekatan umum adalah dengan menggunakan uji deviance ( $G^2$ ) yang membandingkan nilai deviance model penuh dengan deviance model null. Nilai  $G^2$  dihitung dari perbedaan log-likelihood antara kedua model dan mengikuti distribusi Chi-square. Jika nilai  $G^2$  lebih besar dari nilai kritis Chi-square pada derajat kebebasan tertentu (jumlah prediktor), maka dapat disimpulkan bahwa model secara keseluruhan signifikan (Nattino, Pennell, & Lemeshow, 2020).

## 2.3 LINEAR DISKRIMINAN ANALYSIS

Linear Diskriminan Analisis dikembangkan oleh Ronald A. Fisher pada tahun 1933 sebagai metode statistik untuk membedakan dua atau lebih kelompok berdasarkan variabel numerik. Teknik ini membentuk kombinasi linier dari variabel-variabel prediktor untuk memaksimalkan separasi antar kelompok. Rumus dasar LDA adalah:

$$\text{Score} = \sum_{i=0}^p a_i X_i$$

(1)

$$\text{Score} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_p X_p \quad (2)$$

Dengan  $a_i$  sebagai koefisien (bobot) dari setiap variabel  $X_i$ . Studi Altman (1968) dianggap sebagai referensi yang

digunakan oleh LDA untuk mengembangkan model prediksi kebangkrutan perusahaan berdasarkan lima rasio keuangan.

### 2.3.1 Uji Asumsi Analisis Diskriminan

Agar fungsi diskriminan yang dihasilkan optimal, data harus memenuhi beberapa asumsi, yakni distribusi normal multivariat, kesamaan matriks varians-kovarians antar kelompok, homogenitas ragam peragam, dan vektor nilai rata-rata kedua kelompok.

#### 1. Pengujian Normal Multivariat

Menurut Johnson dan Wichern (2007), normalitas multivariat dapat diuji melalui grafik distribusi nilai  $d_j^2$  dengan Chi-Square  $\left[\frac{j-0,5}{n}\right]$ . Jika titik-titik mendekati garis lurus, maka data dianggap mengikuti distribusi normal multivariat Statistik uji yang digunakan :

$$d_j^2 = [X_j - \bar{X}] S^{-1} [X_j - \bar{X}]^T, \quad j = 1, 2, 3, \dots, n$$

Hipotesis yang digunakan adalah

H0 : data berdistribusi multivariat normal.

H1 : data tidak berdistribusi multivariat normal.

#### 2. Uji Kesamaan Matriks Kovarians (Uji Box's M)

Menurut Morisson (1967) untuk uji kehomogenan matriks kovarian dapat dilakukan dengan uji Box's M. Uji Box's M dapat dirumuskan sebagai berikut :

$$S = \frac{1}{\sum_{i=1}^g n_i} \sum_{i=1}^g n_i S_i$$

$$M = \sum_{i=1}^g n_i \ln \ln |S| - \sum_{i=1}^g n_i \ln \ln |S_i|$$

$$C^{-1} = 1 - \frac{2p^3 + 3p - 1}{6(p+1)(g-1)} \left[ \sum_{i=1}^g \frac{1}{n_i} - \frac{1}{\sum_{i=1}^g n_i} \right]$$

dengan hipotesis sebagai berikut

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k = \Sigma$$

$$H_1 : \text{minimal ada satu } \Sigma_i \neq \Sigma$$

$$i = 1, 2, \dots, k$$

Daerah penolakan untuk hipotesis nol dapat dihipotesis dengan menggunakan sebaran *chi-square* sebagai berikut:

$$MC^{-1} > X^2_{a, db}$$

### 3. Uji Homogenitas Matriks Ragam Peragam (Uji Bartlett)

Uji Bartlett digunakan untuk menguji apakah varians dan kovarians antar kelompok adalah seragam, yang merupakan asumsi penting dalam analisis diskriminan. Jika uji ini menunjukkan bahwa varians antar kelompok tidak homogen, maka asumsi kesamaan kovarians tidak terpenuhi yang dapat mengganggu keakuratan hasil analisis diskriminan.

### 4. Uji Vektor Nilai Rata – Rata Kedua Kelompok (Wilk's Lambda)

Wilks' Lambda adalah statistik uji yang digunakan untuk mengukur sejauh mana variabel independen dapat membedakan

antara dua atau lebih kelompok dalam analisis diskriminan. Wilks' Lambda menghitung rasio antara dalam kelompok kovarians dan total kovarians, di mana semakin kecil nilai Wilks' Lambda, semakin besar perbedaan antara kelompok.

### 2.3.2 Metode untuk Membentuk Fungsi Diskriminan

Dalam penyusunan fungsi diskriminan ada dua metode yang dapat digunakan, yaitu estimasi simultan dan estimasi stepwise. Untuk mengetahui variabel bebas yang dapat mendeskripsikan suatu kelompok, digunakan statistik uji Wilks' Lamda. Wilks' Lamda dihitung menggunakan rumus :

$$\Delta^* = \frac{\left| \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \right|}{\left| \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})' \right|}$$

### 2.3.3 Fungsi Diskriminan

$$F \delta_k(X) = X^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k) \quad k(X)$$

merepresentasikan skor keanggotaan suatu observasi terhadap kelompok ke-k. Fungsi ini mempertimbangkan probabilitas awal (prior probability) dan digunakan untuk mengklasifikasikan data ke dalam kelompok dengan nilai  $\delta_k$  tertinggi. Dalam konteks kebangkrutan, fungsi ini membantu membedakan perusahaan bangkrut dan tidak bangkrut serta memberi skor yang mencerminkan

seberapa kuat suatu pengamatan termasuk ke dalam kelompok tertentu.

#### 2.3.4 Kriteria Perbandingan

Kriteria perbandingan teknik klasifikasi didasarkan pada kesalahan klasifikasinya yang dikenal dengan Apparent Error Rate (APER) merupakan nilai dari besar kecilnya jumlah observasi yang salah dalam pengklasifikasian berdasarkan suatu fungsi klasifikasi (Johnson & Wichern 2007). Adapun APER dihitung dengan terlebih dahulu membuat tabel klasifikasi sebagai berikut:

Tabel 2.2 Klasifikasi *actual* dan *predicted* group

Actual group	<u>Predicted group</u>	
	$\pi_0$	$\pi_1$
$\pi_0$	$n_{00}$	$n_{01} = n_0 - n_{00}$
$\pi_1$	$n_{10} = n_1 - n_{11}$	$n_{11}$

Sumber : Johnson and Wichern (2007)

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1}$$

Tingkat keberhasilan klasifikasi dihitung menggunakan *Hit Ratio*:

$$HitRatio = \frac{n_{00} + n_{11}}{n_0 + n_1} \times 100\%$$

## 2.4 Evaluasi

### 2.4.1 Akurasi

Akurasi menunjukkan proporsi klasifikasi yang benar terhadap total data. Gardenia (2015) menyatakan bahwa semakin tinggi kesamaan antara



hasil klasifikasi model dan data aktual, maka semakin baik akurasi model tersebut. Rumus akurasi:

$$Akurasi = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah data}} \times 100\%$$

#### 2.4.2 Confusion Matriks

Confusion matrix adalah alat evaluasi performa model klasifikasi yang menyajikan informasi tentang prediksi model dibandingkan dengan label aktual. Tabel ini terdiri dari empat elemen utama: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Dari keempat elemen ini dapat dihitung berbagai metrik evaluasi seperti akurasi, sensitivitas (recall), dan spesifisitas. Rumus-rumus umum yang digunakan adalah sebagai berikut:

- Akurasi =  $(TP + TN) / (TP + TN + FP + FN)$
- Sensitivitas (Recall) =  $TP / (TP + FN)$
- Spesifisitas =  $TN / (TN + FP)$

Confusion matrix sangat penting dalam analisis regresi logistik dan machine learning karena membantu mengidentifikasi ketidakseimbangan kelas atau bias prediksi, yang tidak dapat ditangkap hanya dengan nilai akurasi saja. Dalam kasus data yang tidak seimbang, sensitivitas dan spesifisitas sering

kali lebih informatif dibandingkan akurasi (Fawcett, 2020).

#### 2.4.3 ROC dan AUC

Receiver Operating Characteristic (ROC) curve adalah grafik yang menggambarkan kinerja model klasifikasi pada berbagai ambang keputusan. Grafik ROC memplot *True Positive Rate* (TPR atau sensitivitas) terhadap *False Positive Rate* (FPR =  $1 - \text{spesifisitas}$ ) untuk setiap nilai threshold yang mungkin. ROC membantu mengevaluasi trade-off antara sensitivitas dan spesifisitas.

AUC (Area Under Curve) adalah luas di bawah kurva ROC, yang menunjukkan probabilitas bahwa model akan memberikan skor yang lebih tinggi untuk kasus positif dibandingkan negatif. AUC memiliki nilai antara 0 dan 1. Semakin mendekati 1, semakin baik model dalam membedakan antara kelas. Model yang memiliki AUC sebesar 0.5 dianggap tidak lebih baik dari tebak-tebakan acak. Menurut Huang et al. (2020), AUC adalah ukuran robust karena tidak bergantung pada threshold tertentu dan bekerja baik untuk data tidak seimbang.

### 3. METODE PENELITIAN

#### 3.1 Jenis dan Sumber Data

Dalam penelitian ini data yang digunakan adalah data yang digunakan berasal dari Kaggle berjudul “*Bankruptcy Italian Companies 2023*” merupakan kumpulan data yang berasal dari Italian Companies pada tahun 2023 dengan jumlah 6.819 data. Setiap entri dalam dataset mewakili satu perusahaan dengan nilai-nilai numerik dari berbagai rasio dan indikator keuangan yang mencerminkan kondisi dan performa keuangan perusahaan tersebut pada suatu waktu. Setiap atribut bertipe numerik dan tidak ada atribut kategorikal, yang membuat dataset ini sangat cocok untuk klasifikasi menggunakan regresi logistik dan analisis diskriminan.

Dataset ini memiliki 96 kolom (atribut) secara keseluruhan, termasuk satu kolom target bernama *Bankrupt?* yang memiliki nilai biner (1 untuk bangkrut, 0 untuk tidak bangkrut). Atribut lainnya terdiri dari 95 fitur numerik yang merupakan indikator dan rasio keuangan yang digunakan dalam analisis kinerja perusahaan. Fitur-fitur ini mencerminkan aspek-aspek penting dari keuangan

perusahaan, seperti profitabilitas, likuiditas, efisiensi operasional, struktur modal, dan pertumbuhan laba.

#### 3.2 Variabel Penelitian

Variabel Respon(Y) = *Bankrupt*

Penelitian ini menggunakan satu variabel respon sebagai berikut :

0 = Kondisi Perusahaan tidak bangkrut

1 = Kondisi perusahaan bangkrut

Variabel Bebas (X) : *ROA(C) before interest and depreciation before interest* ( $X_1$ ), *ROA(A) before interest and % after tax* ( $X_2$ ), *ROA(B) before interest and depreciation after tax* ( $X_3$ ), *Operating Gross Margin* ( $X_4$ ), *Realized Sales Gross Margin* ( $X_5$ ), *Operating Profit Rate* ( $X_6$ ), *Pre-tax net Interest Rate* ( $X_7$ ), *After-tax net Interest Rate* ( $X_8$ ), *Non-industry income and expenditure/revenue* ( $X_9$ ), *Continuous interest rate (after tax)* ( $X_{10}$ ), *Operating Expense Rate* ( $X_{11}$ ), *Research and development expense rate* ( $X_{12}$ ), *Cash flow rate* ( $X_{13}$ ), *Interest-bearing debt interest rate* ( $X_{14}$ ), *Tax rate (A)* ( $X_{15}$ ), *Net Value Per Share (B)* ( $X_{16}$ ), *Net Value Per Share (A)* ( $X_{17}$ ), *Net Value Per Share(C)* ( $X_{18}$ ), *Persistent EPS in the Last Four Seasons* ( $X_{19}$ ), *Cash Flow Per Share* ( $X_{20}$ ), *Revenue Per Share* ( $X_{21}$ ), *Operating Profit Per Share* ( $X_{22}$ ), *Per*

*Share Net profit before tax* ( $X_{23}$ ),  
*Realized Sales Gross Profit Growth Rate*  
( $X_{24}$ ), *Operating Profit Growth Rate*  
( $X_{25}$ ), *After-tax Net Profit Growth Rate*  
( $X_{27}$ ), *Regular Net Profit Growth Rate*  
( $X_{28}$ ), *Continuous Net Profit Growth*  
*Rate* ( $X_{29}$ ), *Total Asset Growth Rate*  
( $X_{30}$ ), *Net Value Growth Rate* ( $X_{31}$ ), *Total*  
*Asset Return Growth Rate Ratio* ( $X_{32}$ ),  
*Cash Reinvestment %* ( $X_{33}$ ), *Current*  
*Ratio* ( $X_{34}$ ), *Quick Ratio* ( $X_{35}$ ), *Interest*  
*Expense Ratio* ( $X_{36}$ ), *Total debt/Total net*  
*worth* ( $X_{37}$ ), *Debt ratio %* ( $X_{38}$ ), *Net*  
*worth/Assets* ( $X_{39}$ ), *Long-term fund*  
*suitability ratio (A)* ( $X_{40}$ ), *Borrowing*  
*dependency* ( $X_{41}$ ), *Contingent*  
*liabilities/Net worth* ( $X_{42}$ ), *Operating*  
*profit/Paid-in capital* ( $X_{43}$ ), *Inventory*  
*and accounts receivable/Net value* ( $X_{44}$ ),  
*Total Asset Turnover* ( $X_{45}$ ), *Accounts*  
*Receivable Turnover* ( $X_{46}$ ), *Average*  
*Collection Days* ( $X_{47}$ ), *Inventory*  
*Turnover Rate (times)* ( $X_{48}$ ), *Fixed Assets*  
*Turnover Frequency* ( $X_{49}$ ), *Net Worth*  
*Turnover Rate (times)* ( $X_{50}$ ), *Revenue*  
*per person* ( $X_{51}$ ), *Operating profit per*  
*person* ( $X_{52}$ ), *Allocation rate per person*  
( $X_{53}$ ), *Working Capital to Total Assets*  
( $X_{54}$ ), *Quick Assets/Total Assets* ( $X_{55}$ ),  
*Current Assets/Total Assets* ( $X_{56}$ ),  
*Cash/Total Assets* ( $X_{57}$ ), *Quick*

*Assets/Current Liability* ( $X_{58}$ ),  
*Cash/Current Liability* ( $X_{59}$ ), *Current*  
*Liability to Assets* ( $X_{60}$ ), *Operating*  
*Funds to Liability* ( $X_{61}$ ),  
*Inventory/Working Capital* ( $X_{62}$ ),  
*Inventory/Current Liability* ( $X_{63}$ ),  
*Current Liabilities/Liability*( $X_{64}$ ),  
*Working Capital/Equity* ( $X_{65}$ ), *Current*  
*Liabilities/Equity* ( $X_{66}$ ), *Long-term*  
*Liability to Current Assets* ( $X_{67}$ ),  
*Retained Earnings to Total Assets* ( $X_{68}$ ),  
*Total income/Total expense* ( $X_{69}$ ), *Total*  
*expense/Assets* ( $X_{70}$ ), *Current Asset*  
*Turnover Rate* ( $X_{71}$ ), *Quick Asset*  
*Turnover Rate* ( $X_{72}$ ), *Working capital*  
*Turnover Rate* ( $X_{73}$ ), *Cash Turnover Rate*  
( $X_{74}$ ), *Cash Flow to Sales* ( $X_{75}$ ), *Fixed*  
*Assets to Assets* ( $X_{76}$ ), *Current Liability*  
*to Liability* ( $X_{77}$ ), *Current Liability to*  
*Equity* ( $X_{78}$ ), *Equity to Long-term*  
*Liability* ( $X_{79}$ ), *Cash Flow to Total Assets*  
( $X_{80}$ ), *Cash Flow to Liability* ( $X_{81}$ ), *CFO*  
*to Assets* ( $X_{82}$ ), *Cash Flow to Equity*  
( $X_{83}$ ), *Current Liability to Current*  
*Assets* ( $X_{84}$ ), *Liability-Assets Flag* ( $X_{85}$ ),  
*Net Income to Total Assets* ( $X_{86}$ ), *Total*  
*assets to GNP price* ( $X_{87}$ ), *No-credit*  
*Interval* ( $X_{88}$ ), *Gross Profit to Sales*  
( $X_{89}$ ), *Net Income to Stockholder's*  
*Equity* ( $X_{90}$ ), *Liability to Equity* ( $X_{91}$ ),  
*Degree of Financial Leverage (DFL)*

( $X_{92}$ ), *Interest Coverage Ratio (Interest expense to EBIT)* ( $X_{93}$ ), *Net Income Flag* ( $X_{94}$ ), *Equity to Liability* ( $X_{95}$ )

### 3.3 Prosedur Penelitian

Pemisahan dataset menjadi dua bagian, yaitu data learning dan data testing. Komposisi pembagian data learning dan data testing pada penelitian ini adalah 70% dan 30%. Langkah - Langkah yang dilakukan pada analisis diskriminan linear :

1. Menghapus Kolom Non-numerik dan Mengubah Semua Menjadi Kolom Bertipe Numerik, Menghapus Kolom Dengan Missing Value dan Standar Deviasi yang Bernilai 0, Standarisasi dan Reduksi Dimensi (PCA).
2. Menguji Asumsi ( Uji asumsi multivariat normal, uji asumsi matriks kovarian dari semua variabel independen sama, Uji homogenitas matriks ragam peragam, Uji vektor nilai rata – rata kedua kelompok)
3. Split data dan training data
4. Membentuk fungsi diskriminan Menentukan nilai batas pemisah kedua kelompok.

5. Interpretasi koefisien dan fungsi diskriminan.
6. Uji signifikansi dilakukan untuk menganalisa model fungsi diskriminan yang terbentuk apakah cukup signifikan atau tidak.
7. Evaluasi menggunakan confusion matrix, ROC curve dan nilai AUC.

#### 3.3.2 Langkah - Langkah yang dilakukan pada analisis Regresi Logistik :

1. Menghapus Kolom Non-numerik dan Mengubah Semua Menjadi Kolom Bertipe Numerik, Menghapus Kolom Dengan Missing Value dan Standar Deviasi yang Bernilai 0, Standarisasi dan Reduksi Dimensi (PCA).
2. Menguji Asumsi ( Uji asumsi multikolinearitas)
3. Split data dan training data
4. Membuat model regresi logistik dan interpretasi koefisien model terhadap variabel independen.
5. Uji signifikansi model regresi logistik
6. Menghitung R-Square

7. Menghitung odds ratio dan menginterpretasikannya
8. Evaluasi menggunakan confusion matrix, ROC curve dan nilai AUC.

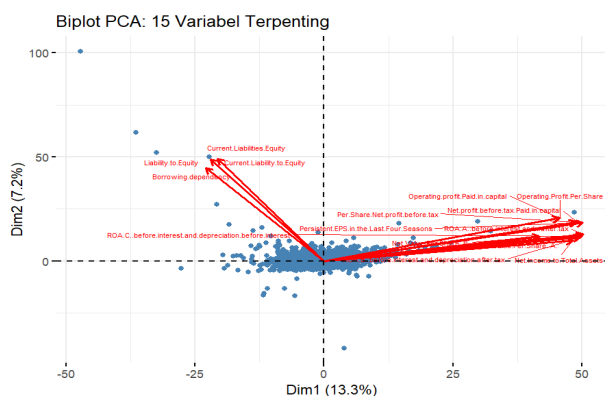
### 3.3.3 Langkah- Langkah Untuk Mencapai Tujuan Ketiga Yaitu

Langkah yang digunakan adalah membandingkan performa model klasifikasi *Logistic Regression* dan Linear Diskriminan Analisis dalam memprediksi kebangkrutan perusahaan serta Metode mana yang lebih optimal digunakan berdasarkan evaluasi akurasi klasifikasi, sensitivitas, dan spesifisitas.

## 4. PEMBAHASAN DAN PENERAPAN

### 4.1 REDUKSI DIMENSI (PCA)

#### 4.1.1 Visualisasi Biplot PCA

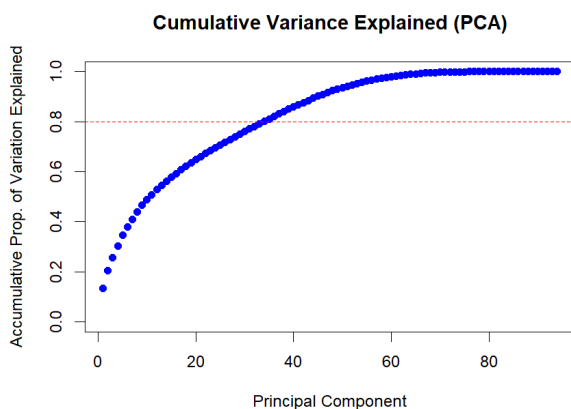


Biplot PCA di atas menunjukkan proyeksi data perusahaan terhadap dua komponen utama (Dim1 dan Dim2) yang secara bersama menjelaskan sekitar 20,5% variabilitas total dalam data (Dim1: 13.3%, Dim2: 7.2%). Titik-titik biru mewakili observasi perusahaan, sementara panah merah mewakili 15 variabel paling berkontribusi terhadap pembentukan dua komponen utama tersebut.

Panah yang lebih panjang dan menjauh dari pusat menunjukkan variabel yang memiliki kontribusi tinggi terhadap dimensi. Misalnya, “Operating Profit Per Share”, “Operating profit.Paid in capital”, dan “Net profit before tax.Paid in capital” sangat berkontribusi positif terhadap Dimensi 1, menunjukkan bahwa komponen ini merepresentasikan kinerja keuangan berbasis profitabilitas. Di sisi lain, variabel seperti “Current Liabilities/Equity” dan “Liability to Equity” memiliki arah berlawanan dan panjang panah yang besar, menandakan kontribusi tinggi tetapi berlawanan arah dengan profitabilitas, cenderung mewakili risiko atau leverage keuangan. Secara keseluruhan, biplot ini memberi gambaran bahwa komponen utama pertama (Dim1) banyak dipengaruhi oleh

indikator profitabilitas vs leverage, dan dapat digunakan untuk memisahkan perusahaan berdasarkan kesehatan finansialnya. Sedangkan Dimensi 2 kurang dominan tetapi tetap mencerminkan beberapa variasi tambahan, seperti rasio hutang dan efisiensi.

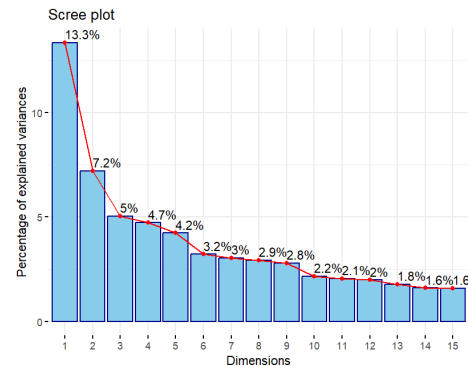
#### 4.1.2 Visualisasi PCA



Plot tersebut menunjukkan *Cumulative Variance Explained* dari hasil analisis Principal Component Analysis (PCA). Terlihat bahwa grafik memiliki pola melengkung yang menanjak tajam pada awalnya, lalu melandai setelah sekitar 30–40 komponen. Garis horizontal putus-putus merah menunjukkan ambang

batas sebesar 80% dari total variasi yang ingin dijelaskan. Dari plot ini dapat disimpulkan bahwa sekitar 35 komponen utama sudah cukup untuk menjelaskan sekitar 80% variasi dalam data.

#### 4.1.3 ScatterPlot



Berdasarkan grafik Scatterplot di atas, terlihat bahwa kontribusi varian terbesar berasal dari komponen pertama hingga komponen kesepuluh. Setelah komponen ke-10, tambahan varian yang dijelaskan oleh komponen-komponen berikutnya semakin kecil dan cenderung mendatar, yang menunjukkan adanya *elbow* atau titik siku di sekitar PC ke-10. Meskipun secara kumulatif 35 komponen utama (PC35) mampu menjelaskan sekitar 80% total variasi data, dalam konteks analisis ini kami memilih untuk menggunakan hanya 10 komponen utama pertama (PC10) dengan kumulatif variance 48.5% . Keputusan ini diambil untuk mempermudah proses analisis dan interpretasi, tanpa mengorbankan terlalu

banyak informasi penting dari data.

Pendekatan ini juga didukung oleh hasil Scree Test, di mana penurunan kontribusi varians mulai melambat secara signifikan setelah PC10, menandakan bahwa sebagian besar informasi utama sudah tercakup dalam sepuluh komponen tersebut.

## 4.1.4 Hasil PCA

```
> pca_resultssvd$V
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.2388067650 0.0770361680 -3.868752e-02 -0.076114243 0.012554268
[2,] 0.2398053854 0.0827832645 -2.362209e-02 -0.072477399 0.0103539275
[3,] 0.2369341852 0.0735260387 -4.548348e-02 -0.077309979 0.0144903699
[4,] 0.1212997865 -0.0175762729 -9.782201e-02 0.176611371 -0.0378187786
[5,] 0.1208930839 -0.0178554869 -9.772145e-02 0.177225941 -0.0380029984
[6,] 0.0132100118 0.0181565683 4.169900e-02 0.066469552 0.459086592
[7,] 0.0203703912 0.0167010045 3.644471e-02 0.086577060 0.4851245121
[8,] 0.0200274863 0.0147694367 3.449435e-02 0.091446502 0.4789792922
[9,] 0.0088239879 -0.0110151189 -2.825156e-02 0.012332733 -0.1420905323
[10,] 0.0194901400 0.0159618670 3.451022e-02 0.087627636 0.4848265666
[11,] 0.0064738599 0.0656243887 6.610217e-02 -0.124424809 0.0283417926
[12,] 0.0271398037 -0.0020169816 -9.671899e-03 -0.023659130 0.0157566688
[13,] 0.1200395031 -0.0349798410 -1.710383e-01 0.108570722 -0.0004836220
[14,] 0.0159486793 0.0034988249 3.980957e-03 0.006096921 -0.0031959433
[15,] 0.0619544767 0.0488097784 4.912126e-02 -0.035315138 0.0079320931
[16,] 0.1990098524 0.0763464719 -3.179098e-02 -0.116905020 0.0064408924
[17,] 0.1990729186 0.076043457 -3.153907e-02 -0.116757132 0.0063997716
[18,] 0.1990157331 0.0760944004 -3.149526e-02 -0.116843920 0.0064280610
[19,] 0.2394792624 0.1185323043 1.925878e-02 -0.103009592 -0.0079923081
[20,] 0.1422409203 0.0399813776 -7.934597e-02 -0.017009381 -0.0017583930
[21,] 0.0073834178 -0.0170451719 -4.872488e-02 0.063795123 -0.0114572581
[22,] 0.2181685568 0.1337609516 5.566299e-02 -0.070303869 -0.0179847027
[23,] 0.2328848574 0.1213436929 2.721752e-02 -0.103198118 -0.0083366257
[24,] -0.0031978921 0.0072841082 2.739072e-02 0.011427523 -0.0043957572
[25,] 0.0137375815 0.0155915443 6.821899e-03 -0.002782706 0.0108129175
[26,] 0.0327041808 0.0221024300 4.743863e-03 -0.014131075 0.0246234904
[27,] 0.0326066377 0.0221056691 4.862313e-03 -0.014489413 0.0245994659
[28,] 0.0091583703 0.0013355476 -4.280078e-03 -0.007024454 0.0039615667
[29,] 0.0044907838 -0.0249573756 -4.780156e-02 -0.028034008 0.0395143877

[30,] -0.0112727044 -0.0279153746 4.527668e-02 -0.029861678 0.0010501148
[31,] 0.0197766395 0.0188218217 2.270469e-02 -0.009012744 0.0025484561
[32,] 0.0995323957 -0.0248961889 -6.183264e-02 0.003633092 0.0085870955
[33,] 0.0008420309 0.0008913198 -1.090514e-02 -0.024508075 -0.0029343227
[34,] -0.0060842396 0.0073105438 7.120450e-03 -0.007905819 -0.0062176321
[35,] -0.0016548441 0.0039781318 -3.758597e-03 -0.015308063 0.0051786481
[36,] 0.0105001921 -0.0228446387 -4.811596e-02 0.058356569 -0.0068809479
[37,] -0.1308537956 0.1818192944 1.673310e-01 -0.161526232 0.0206354096
[38,] 0.1308537956 -0.1818192944 -1.673310e-01 0.161526232 0.0206354096
[39,] 0.0141026006 0.0172358865 3.227828e-02 0.036338223 -0.0155228500
[40,] -0.1095566117 0.2894568754 -1.806799e-01 0.072568336 -0.0010439718
[41,] -0.0504845725 0.1860779918 -1.786466e-01 0.066517891 -0.0021039917
[42,] 0.2174153033 0.1338360978 5.523516e-02 -0.069961596 -0.0180668843
[43,] 0.2341555616 0.1215728221 2.401094e-02 -0.101532261 -0.0082766009
[44,] -0.0699607792 0.2630334609 5.350000e-02 0.069764831 -0.0158807552
[45,] 0.0461966813 0.1638327512 2.639844e-01 0.026317303 -0.0213424215
[46,] -0.0069095087 0.0044199620 -1.640836e-02 0.037498455 0.0248212103
[47,] 0.0009339184 0.005081741 -9.50552e-05 -0.006175884 -0.0015803286
[48,] -0.0155228450 -0.0348796201 -3.404804e-02 0.047785743 0.0032231516
[49,] -0.0374855664 -0.0697632241 -1.950537e-01 -0.090539005 0.0428788022
[50,] -0.0154401744 0.1936967614 2.229498e-01 -0.006217090 -0.0159884768
[51,] -0.0084097962 0.0013972739 -1.920505e-02 -0.010779580 -0.0536797820
[52,] 0.0972605152 0.0785866904 3.509551e-02 -0.026691462 -0.0072871549
[53,] -0.0018171303 0.0058271127 2.166301e-02 0.014274434 -0.0049544809
[54,] 0.1357046544 -0.0344759277 8.012762e-02 0.291178029 -0.0524033115
[55,] 0.0909915253 0.069277710 2.158930e-01 0.265195302 -0.0682206127
[56,] 0.0620732319 0.1140461668 2.851736e-01 0.250659688 -0.0653933691
[57,] 0.1201006993 -0.0171666475 3.398969e-02 0.268899125 -0.0624059592
[58,] -0.0040092474 -0.0138018580 4.561443e-02 -0.031002904 -0.0022656344
[59,] -0.0177065474 0.0130536911 4.354333e-01 -0.022041448 0.0045687846
[60,] -0.0959854376 0.1972644241 2.742881e-01 -0.049514741 -0.0180990808
```

```
[61,] 0.1370825504 -0.0269146977 -1.482475e-01 0.119155812 -0.0088827388
[62,] -0.0033227577 -0.001815407 -4.581542e-03 -0.009464396 0.0099018527
[63,] 0.0036378034 0.0067338170 -5.784163e-03 -0.004348465 -0.0113516699
[64,] 0.0408346088 0.0726253829 2.336559e-01 0.201678434 -0.0690455570
[65,] 0.0812374924 -0.1748639457 2.487290e-01 -0.009188169 -0.0092056677
[66,] -0.0982966394 0.3193930798 -1.338861e-01 0.101458567 -0.0149555856
[67,] 0.0081832723 0.0090433152 1.474393e-02 0.012352033 -0.0046557823
[68,] 0.1817749734 0.0527091820 -4.891800e-02 -0.099732642 0.0233103882
[69,] 0.0094347119 0.0007170045 -1.403891e-02 -0.008238424 -0.0014461436
[70,] -0.0619117233 -0.0095768738 9.083715e-02 0.215633064 -0.0525836973
[71,] -0.0162020597 0.0545734231 5.228281e-02 -0.108843646 0.0242815283
[72,] -0.0327987957 0.0620329003 5.305451e-02 -0.133744545 0.0348019876
[73,] 0.0093120278 -0.0119299029 -1.890534e-02 0.060540110 0.0374009883
[74,] -0.0275733627 0.0235551105 3.969335e-02 -0.097666928 0.0314692026
[75,] 0.0084905435 -0.0043372123 -2.706318e-03 0.054765793 0.1113325136
[76,] -0.0036248971 0.0069380501 1.023349e-02 0.06401065 -0.0031230544
[77,] 0.0408346088 0.0726253829 2.336559e-01 0.201678434 -0.0690455570
[78,] -0.0982966394 0.3193930798 -1.338861e-01 0.101458567 -0.0149555856
[79,] -0.0836077976 0.2191883795 -1.885450e-01 0.027506336 0.0092299995
[80,] 0.0990336187 0.0320603625 1.070453e-02 0.111569569 -0.0093711751
[81,] 0.0705729340 0.0078321374 -2.222844e-02 0.127013593 -0.0206764428
[82,] 0.1549460381 -0.0017637995 -1.334170e-01 0.027804606 0.0093476392
[83,] 0.0666346024 -2.477090e-01 9.803254e-02 0.185705704 -0.00837502
[84,] -0.0939023345 0.0457239103 8.529328e-04 -0.199418744 -0.0437855596
[85,] -0.0266387613 -0.0761354438 1.265813e-01 -0.054785644 -0.0012258474
[86,] 0.2301665243 0.0633602265 3.289401e-02 -0.078222442 0.0147478732
[87,] -0.0219771978 0.0044045760 3.204137e-02 0.040173348 -0.0127910308
[88,] 0.0066656486 -0.0070287455 -1.492108e-02 0.017240164 -0.0022701102
[89,] 0.1213001901 -0.0175759584 -9.78211e-02 0.176611201 -0.0378182874
[90,] 0.1146822204 -0.2071909127 1.839287e-01 -0.134613301 0.0140019600
[91,] -0.1046362124 0.3174027087 -1.665979e-01 0.085341649 -0.0077270743
```

```
[91,] -0.1046362124 0.3174027087 -1.665979e-01 0.085341649 -0.0077270743
[92,] -0.0071492672 -0.0031043837 -7.298980e-03 -0.014074045 0.0054713635
[93,] 0.0014556807 0.0059661921 7.232073e-03 -0.005238927 0.0005098628
[94,] 0.0658662203 -0.1253414430 -1.297066e-01 0.190926797 -0.0347774736

      [,6]      [,7]      [,8]      [,9]      [,10]
[1,] 0.0361559580 -1.216318e-02 0.0413438717 0.0217716677 -0.140497028
[2,] -0.0056316051 -7.119305e-03 0.0334230489 0.0251798165 -0.167186866
[3,] 0.0309222354 -1.059201e-02 0.0410566009 0.0253278131 -0.151284261
[4,] 0.0666346024 -2.477090e-01 9.803254e-02 0.185705704 -0.00837502
[5,] 0.0860910749 -2.476997e-01 0.3071508318 -0.1855187457 -0.045576373
[6,] 0.0387285354 -1.707181e-01 -0.1062033608 0.0301601175 0.025312584
[7,] -0.0012400462 -5.752818e-03 0.013839533 0.0303531859 0.01496814
[8,] -0.0169161755 6.318449e-02 0.0495558183 -0.0532853918 0.009201159
[9,] -0.0969093475 4.045872e-01 0.038570248 -0.1346570248 -0.033113334
[10,] 0.0032694673 1.155704e-03 0.0112600484 -0.0422530575 0.016835625
[11,] 0.0022967680 1.037570e-01 0.0757378774 0.0686677615 -0.079389907
[12,] -0.023283274 1.046399e-02 0.0274425612 0.0296257054 -0.009337075
[13,] 0.2633503713 5.661571e-02 0.0169565131 0.0829265085 -0.048032463
[14,] -0.0241976989 -6.431885e-03 0.0088517932 -0.0022858624 -0.001310148
[15,] 0.0069343230 -1.640065e-03 0.0245376727 0.0247361955 -0.169509026
[16,] -0.1611420896 8.623763e-03 0.0967193725 -0.0743330838 0.208149938
[17,] -0.1613515539 8.587686e-03 -0.0962512710 -0.0743074068 0.207431089
[18,] -0.1613504782 8.528547e-03 -0.0961185747 -0.0744050734 0.207504791
[19,] -0.0599814532 -2.385127e-02 -0.0081829642 -0.0727560011 0.110439606
[20,] 0.2423594165 6.420537e-03 -0.0636148883 0.0307113625 0.055251666
[21,] 0.0577544667 7.738116e-02 0.0751791121 -0.0952443203 0.040758232
[22,] 0.0007913919 -3.572269e-02 0.0189178536 -0.0781709319 0.101271400
[23,] -0.0612276492 -2.533185e-02 -0.0009067983 -0.0670824765 0.107633790
[24,] -0.0191159599 -1.412537e-02 0.0224006383 -0.0058415224 -0.021328880
[25,] -0.1176591283 -5.59978e-03 0.2484853862 0.3896594835 0.094334144
[26,] -0.1395091546 -5.498913e-02 0.3056931777 0.4464988940 0.094085053
```

```
[27,] -0.1400065100 -5.466682e-02 0.3048265349 0.4463534196 0.093675342
[28,] -0.0428993088 -1.23632e-02 0.0499155939 0.0819871865 0.039266947
[29,] 0.0115410020 -6.076959e-03 0.0092583800 0.0259115755 -0.195473885
[30,] 0.0832292976 -8.261447e-03 0.0415231131 -0.0195834674 0.106400480
[31,] -0.0950903871 -4.079972e-02 0.1354196618 0.1748177126 0.090025721
[32,] 0.3107984493 8.394678e-02 -0.0231592994 0.1195432550 0.097776908
[33,] 0.0298015861 -3.643781e-02 0.0276347548 -0.0191421697 0.024191693
[34,] -0.0069588874 -2.043116e-02 0.0103365697 -0.0189569867 0.028099190
[35,] 0.0125734442 1.317594e-03 0.013424685 0.0073670145 -0.020919006
[36,] 0.0625629458 6.682056e-02 -0.0578076882 0.0718514293 0.031027642
[37,] 0.1843762937 -2.173248e-03 0.1289227759 -0.0565843656 0.098562737
[38,] -0.1843762937 -2.173248e-03 -0.1289227759 0.0565843656 0.098562737
[39,] -0.0333588391 1.805169e-02 -0.0084354925 -0.0239002437 0.053676321
[40,] -0.0255169030 8.763605e-03 -0.0143948279 0.0012109447 0.026662008
[41,] -0.0236468842 1.287511e-03 -0.0398527752 0.0192985920 -0.039018202
[42,] 0.0029046421 -3.517417e-02 0.0179829778 -0.0797590029 0.100067190
[43,] -0.0557421363 -2.273783e-02 -0.0041222433 -0.0697030934 0.097627937
[44,] -0.0240143456 9.56862e-03 0.0177383430 -0.0193665797 -0.040237098
[45,] 0.0551465512 6.829086e-02 -0.0217025075 0.0939902035 -0.178747902
[46,] 0.0111071520 6.277831e-02 0.1145805834 -0.1135738836 0.059217604
[47,] 0.0164539360 -3.130097e-02 0.0369626685 -0.0279553134 -0.010999816
[48,] -0.0316735253 -4.897321e-02 0.0245983314 -0.0300702550 0.017975346
[49,] 0.0604500308 -1.450489e-02 0.0696252002 -0.0098661927 0.115235539
[50,] 0.1444993848 7.026939e-02 0.0217764701 0.0708363462 -0.072786686
[51,] 0.0552313418 -5.790838e-02 -0.0568511573 -0.071446349 0.021575428
[52,] -0.0231935713 7.466135e-02 0.0816168422 -0.0837438974 -0.020384916
[53,] -0.0061260746 7.128554e-03 -0.0042393625 0.0002634215 -0.002481354
[54,] -0.1597347557 2.580614e-02 -0.1073846950 0.0430280883 -0.056960566
[55,] -0.0219469857 2.481393e-02 -0.0717677874 0.0436195469 0.002982143
[56,] -0.0520853711 1.063819e-02 0.083243245 0.0067588852 -0.031627161
[57,] 0.0433964182 4.476031e-02 -0.0966641136 0.0640227050 0.179426828
```



```
[58.] 0.0569062319 -1.954501e-02 0.0418627800 -0.0197518376 0.054330531
[59.] 0.0119422634 -2.807402e-03 0.0176563263 -0.0189154034 0.025447215
[60.] 0.1409212292 -1.980565e-02 0.0904635020 -0.0476438393 0.032867691
[61.] 0.2580289669 8.345692e-02 -0.0457289592 0.1177829002 -0.057555599
[62.] -0.0090680801 2.492837e-02 0.0088742760 -0.0069776639 0.027367818
[63.] 0.0624694280 -2.407758e-02 0.0185706079 -0.0281093226 0.056325697
[64.] -0.0351734216 -4.861440e-02 -0.0557611181 0.0019682344 -0.100978652
[65.] 0.0158392706 1.219891e-02 0.0349541262 -0.0202081226 0.013069116
[66.] -0.0229425169 5.143999e-03 -0.0267390284 0.0082825913 -0.002571348
[67.] 0.0087669342 -2.044156e-05 -0.0079419015 0.0076911444 -0.027567438
[68.] -0.0371493265 4.572626e-02 -0.0415264993 0.0465068472 -0.207439879
[69.] 0.0157740722 -3.362797e-02 0.0335030974 -0.0272820163 -0.003231225
[70.] 0.0463235507 -1.331887e-01 0.1040160016 -0.0958301565 0.095592463
[71.] 0.0938144677 6.372834e-02 0.0026115464 0.0680379839 -0.184774169
[72.] 0.0619413809 3.534487e-02 0.0384786064 0.0364077140 -0.194446569
[73.] -0.0912774177 4.541273e-01 0.2662493483 -0.1638708417 -0.021523844
[74.] -0.0138514167 -2.438320e-02 0.0574685329 -0.0107284134 -0.170382926
[75.] -0.0938564639 4.488167e-01 0.2357270683 -0.1487298840 -0.017079278
[76.] -0.0164220564 -2.967362e-03 0.0068561229 -0.0172467016 0.029002508
[77.] -0.0351734216 -4.861440e-02 -0.0557611181 0.0019682344 -0.100978652
[78.] -0.0229425169 5.143999e-03 -0.0267390284 0.0082825913 -0.002571348
[79.] 0.0183421403 2.199192e-02 -0.0130572838 0.0118279141 0.022808100
[80.] 0.2077011161 1.629078e-01 -0.0531182238 0.1183898079 0.268629155
[81.] 0.2094329466 1.357689e-01 -0.0311347626 0.1045177915 0.230118188
[82.] 0.3062387543 8.129576e-02 -0.0410981094 0.1181106778 -0.104972981
[83.] 0.1787769365 1.241327e-01 -0.0268741703 0.0598738508 0.248300806
[84.] 0.2058823903 -1.248545e-01 0.0721590601 -0.0257966586 0.107211460
[85.] 0.1958719027 -8.671419e-03 0.0706130345 -0.0471049959 0.154656674
[86.] -0.0114752856 7.612510e-03 0.0200216435 0.0429761648 -0.213420326
[87.] -0.0011055924 -2.809354e-02 0.0414242244 -0.0358116742 0.077947648
[88.] -0.0199583665 1.376230e-02 -0.0233417207 0.0113513329 -0.003530530
```

```
[89.] 0.0857076818 -2.477096e-01 0.3072044850 -0.1857069024 -0.045437989
[90.] 0.0494432000 5.226234e-05 0.0488369494 -0.0121973644 -0.067776887
[91.] -0.0103182146 1.067147e-02 -0.0234710484 0.0099013958 0.007214746
[92.] 0.0112467673 1.719272e-03 0.0080253817 0.0045521742 0.003041733
[93.] 0.0083757124 1.755778e-03 0.0069783496 0.0062418136 -0.017982388
[94.] -0.0875539165 5.333810e-02 -0.1215573132 0.0576695285 0.003823698
```

Interpretasi :

### PC1 (13,3%) → Profitabilitas

Variabel dengan nilai loading terbesar:

- Variabel ke-43 = Net profit before tax/Paid-in capital: loading  $\approx 0.234$
- Variabel ke-42 = Operating profit/Paid-in capital: loading  $\approx 0.217$
- Variabel ke-22 = Operating Profit Growth Rate: loading  $\approx 0.218$
- Variabel ke-23 = After-tax Net Profit Growth Rate: loading  $\approx 0.233$
- Variabel ke-19 = Cash Flow Per Share: loading  $\approx 0.239$

### PC2 (7,2%) → Likuiditas & Efisiensi Aset

Variabel dengan loading tinggi:

- Variabel ke-91 = Net Income to Total Assets: loading  $\approx 0.317$

- Variabel ke-90 = Liability-Assets Flag: loading  $\approx 0.219$

- Variabel ke-79 = Net Worth Turnover Rate (times): loading  $\approx 0.220$

- Variabel ke-66 = Quick Assets/Current Liability: loading  $\approx 0.319$

### PC3 (5%) → Efisiensi Keuangan Bersih

Variabel dengan loading dominan:

- Variabel ke-13 = After-tax net Interest Rate: loading  $\approx -0.171$
- Variabel ke-8 = After-tax net Interest Rate: loading  $\approx -0.177$
- Variabel ke-36 = Cash Flow Per Share: loading  $\approx -0.211$

### PC4 (4.7%) → Struktur Modal & Laba Operasi

Variabel dominan:

- Variabel ke-4 = Operating Gross Margin: loading  $\approx 0.176$
- Variabel ke-94 = Equity to Liability: loading  $\approx 0.190$
- Variabel ke-65 = Long-term fund suitability ratio (A): loading  $\approx -0.009$
- Variabel ke-83 = Quick Asset Turnover Rate: loading  $\approx 0.267$

### PC5 (4.2%) → Beban Bunga & Nilai Saham

Variabel dominan:

- Variabel ke-6 = Interest-bearing debt interest rate: loading  $\approx 0.459$



- Variabel ke-7 = Tax rate (A): loading  $\approx 0.485$

- Variabel ke-8 = Net Value Per Share (B): loading  $\approx 0.479$

- Variabel ke-10 = Net Value Per Share (C): loading  $\approx 0.484$

#### **PC6 (3.2%) → Efisiensi Modal & Pertumbuhan Aset**

Variabel dominan:

- Variabel ke-32 = Net Worth Turnover Rate : loading  $\approx 0.311$

- Variabel ke-30 = Total Asset Growth Rate : loading  $\approx 0.258$

- Variabel ke-13 = Cash Flow per Share : loading  $\approx 0.263$

#### **PC7 (3%) → Profitabilitas Saham**

Variabel dominan:

- Variabel ke-20 = Revenue Per Share : loading  $\approx 5.66e-03$

- Variabel ke-29 = Per Share Net profit before tax : loading  $\approx 8.26e-03$

#### **PC8 (2.9%) → Kesehatan Keuangan**

Variabel dominan:

- Variabel ke-93 = Liability-Assets Flag : loading  $\approx 0.308$

- Variabel ke-89 = Net Income to Equity : loading  $\approx 0.307$

#### **PC9 (2.8%) → Leverage & Pertumbuhan Laba**

Variabel dominan:

- Variabel ke-87 = DFL (Degree of Financial Leverage) : loading  $\approx 0.154$

- Variabel ke-24 = Growth of Gross Profit : loading  $\approx 0.0829$

#### **PC10 (2.2%) → Posisi Ekonomi Makro**

Variabel dominan:

- Total Assets to GNP Price (No. 82, Loading: 0.248)

- Net Income to Equity (No. 89, Loading: 0.154)

### **4.2 Analisis Diskriminan Linear**

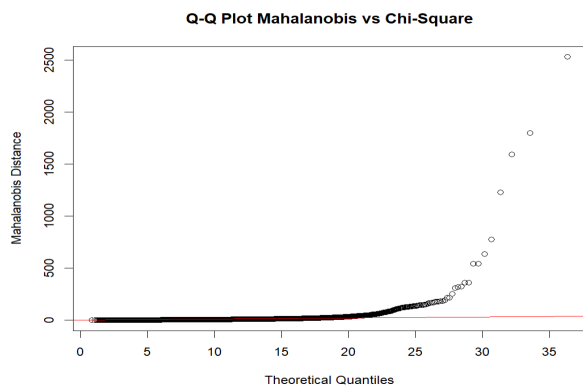
#### **4.2.1 Uji Asumsi Normal Multivariat**

Dalam penelitian ini, pengujian dilakukan terhadap sepuluh komponen utama hasil PCA yang telah distandarisasi dan dibersihkan dari outlier. Pemeriksaan dilakukan menggunakan fungsi `mvn()` dari package *MVN* di R dengan metode Mardia.

```
> print(mardia_result)
$multivariateNormality
      Test      Statistic p value Result
1 Mardia Skewness  922338.48201188      0      NO
2 Mardia Kurtosis  6642.466626056      0      NO
3 MVN              <NA>          <NA>      NO

$univariateNormality
      Test      Variable Statistic      p value Normality
1 Anderson-Darling Dim.1      10.3645 <0.001      NO
2 Anderson-Darling Dim.2      16.1791 <0.001      NO
3 Anderson-Darling Dim.3      14.1708 <0.001      NO
4 Anderson-Darling Dim.4       8.0869 <0.001      NO
5 Anderson-Darling Dim.5      13.9972 <0.001      NO
6 Anderson-Darling Dim.6      30.6362 <0.001      NO
7 Anderson-Darling Dim.7      34.5821 <0.001      NO
8 Anderson-Darling Dim.8      26.1674 <0.001      NO
9 Anderson-Darling Dim.9      45.9137 <0.001      NO
10 Anderson-Darling Dim.10     7.7184 <0.001      NO

$descriptives
      n      Mean      Std.Dev      Median      Min      Max      25th
Dim.1  6785  0.0572243004  2.6237013  0.011451321 -11.355191  12.122131 -1.5517776
Dim.2  6785  0.0367536105  2.3841608  0.197982573 -10.765711  9.129349 -1.3698210
Dim.3  6785  0.0251296665  1.2255349 -0.058801332 -4.469642  7.054362 -0.8018775
Dim.4  6785  0.0006324131  1.5558252  0.017504568 -7.106796  7.960482 -0.9613760
Dim.5  6785  0.0115951429  0.7393480 -0.034313620 -5.715476  3.056771 -0.4647096
Dim.6  6785  0.0208146605  0.4284641 -0.016557094 -3.196953  4.993836 -0.2567644
Dim.7  6785  0.0322929948  0.3795379  0.031949262 -7.652720  2.959952 -0.1759084
Dim.8  6785  0.0110695634  0.7163936  0.040395773 -5.541226  3.563704 -0.3726583
Dim.9  6785  0.0096290551  0.6325183 -0.004458519 -3.769830  7.457024 -0.3355030
Dim.10 6785  0.0122366392  0.2037292  0.005313001 -1.053976  1.691204 -0.1186498
```



Prosedur pendeteksian outlier dilakukan sebelumnya menggunakan Mahalanobis Distance, dengan perbandingan terhadap nilai kritis distribusi Chi-Square (cutoff  $\approx 43.82$  untuk  $df = 10$  dan  $\alpha = 0.001$ ).

Dari grafik Chi-Square Q-Q plot yang dihasilkan, dapat dilihat bahwa sebagian besar titik mengikuti garis linear pada awal sebaran, namun mulai menyimpang pada bagian kuartil atas. Penyimpangan ini menunjukkan bahwa meskipun transformasi dan standardisasi telah dilakukan, serta outlier telah dibuang,

masih terdapat deviasi signifikan pada ekor distribusi. Hal ini mengindikasikan bahwa data belum sepenuhnya memenuhi asumsi normalitas multivariat secara ketat. Sebelumnya telah dilakukan proses identifikasi dan penghapusan outlier menggunakan jarak Mahalanobis dengan cutoff ketat ( $\alpha = 0.0001$ ). Namun, meskipun outlier telah dibuang, pola penyimpangan masih terlihat jelas. Hal ini menunjukkan bahwa data tetap tidak normal secara multivariat, yang juga konsisten dengan hasil uji Mardia yang menghasilkan p-value nol dan hasil “NO” untuk asumsi normalitas. Meski demikian, hasil ini tetap dapat diterima secara praktis mengingat robust-nya metode PCA dan regresi logistik terhadap pelanggaran asumsi ini.

#### 4.2.2 Uji Asumsi Homogenitas Ragam Peragam

```
> uji_bart(data_reduced_no_outlier[, 1:10])

      Bartlett's test of sphericity

data:  data_reduced_no_outlier[, 1:10]
Chi-squared = 68094, df = 45, p-value < 2.2e-16
```

Hasil menunjukkan Nilai p yang sangat kecil ini jauh di bawah ambang signifikansi 0.05, yang berarti hipotesis nol ditolak atau dengan kata lain terdapat perbedaan korelasi ragam-peragam antar kategori bankruptcy.

#### 4.2.3 Pengujian perbedaan rata-rata variabel dependen

Untuk menguji apakah terdapat perbedaan rata-rata ( nilai variabel prediktor) antar kategori bankruptcy, digunakan Uji Manova dengan statistik uji *Wilk's Lambda*. Untuk melakukan uji tersebut di R dapat menggunakan fungsi `manova` dan mengisikan Wilks pada parameter test.

```
> summary(manova_model, test = "wilks")
      Df  wilks approx F num Df den Df    Pr(>F)
Bankrupt 1 0.89098  82.887    10  6774 < 2.2e-16 ***
Residuals 6783
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gambar diatas menunjukkan bahwa dengan tingkat signifikan 5%, didapatkan keputusan untuk menolak hipotesis nol ( $H_0$ ) atau dengan kata lain terdapat perbedaan rata-rata (nilai variabel prediktor) antar kategori bankruptcy.

#### 4.2.4 Kontribusi variabel prediktor terhadap variabel respon

```
> cc

Canonical Discriminant Analysis for Bankrupt:

      CanRsq Eigenvalue Difference Percent Cumulative
1 0.10902    0.12236             100          100

Test of H0: The canonical correlations in the
current row and all that follow are zero

      LR test stat approx F numDF denDF    Pr(> F)
1      0.89098    82.887    10  6774 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hasil Canonical Discriminant Analysis menunjukkan bahwa nilai Canonical R-squared sebesar 0.10902, yang berarti kombinasi linier dari variabel yang digunakan dalam penelitian dapat menjelaskan sebesar 10.90% variasi antara kelompok (perusahaan bangkrut vs tidak bangkrut). Sisanya, yaitu 89.09%, dijelaskan oleh variabel lain di luar model ini. Uji signifikansi menunjukkan bahwa hubungan kanonik ini signifikan ( $p < 2.2e-16$ ), sehingga dapat disimpulkan bahwa model secara statistik mampu membedakan antar kelompok berdasarkan variabel prediktor yang digunakan.

#### 4.2.5 Model Fungsi Linear Diskriminan Analisis

```
> lda_model
Call:
lda(Bankrupt ~ ., data = train_data)

Prior probabilities of groups:
 1
0.96850886 0.03149114

Group means:
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5      Dim.6      Dim.7
0  0.1889596  0.009649349  0.01629535 -0.009527536  0.01730202  0.01733251  0.02895105
1 -4.1118394  0.386321883  0.21870263  0.347861740  0.05391208  0.09011603  0.12929780
      Dim.8      Dim.9      Dim.10
0  0.02453106 -0.01074799  0.01628688
1 -0.22297764  0.35599468 -0.08001068

Coefficients of linear discriminants:
LD1
Dim.1 -0.24783058
Dim.2  0.46350082
Dim.3  1.64381039
Dim.4  1.38395647
Dim.5  0.61950317
Dim.6 -0.03397322
Dim.7  0.53748570
Dim.8 -0.14237134
Dim.9 -0.05241772
Dim.10 1.21090834
```

Berdasarkan hasil output tersebut, menunjukkan bahwa sekitar 97% train data berasal dari perusahaan yang tidak bangkrut, dan hanya sekitar 3% dari

perusahaan bangkrut. Menunjukkan adanya ketidakseimbangan kelas dalam data.

Group means, memperlihatkan rata-rata nilai dari 10 dimensi ( hasil PCA) untuk masing-masing kelompok (0 = tidak bangkrut, 1=bangkrut). Dim.1, nilai rata-rata untuk perusahaan tidak bangkrut adalah 0.1889, sementara untuk yang bangkrut adalah -4.1118. Selisih yang besar ini mengindikasikan Dim.1 sangat berperan dalam membedakan kedua kelas.

Sedangkan koefisien - koefisien dapat digunakan untuk membentuk fungsi diskriminan linear(LD1), yaitu :

$$Y = -0.24783058X_1 + 0.46350082X_2 + 1.64381039X_3 + 1.38395647X_4 + 0.61950317X_5 - 0.03397322X_6 + 0.53748570X_7 - 0.14237134X_8 - 0.05241772X_9 + 1.21090834X_{10}$$

Dimana berdasarkan hasil PCA didapat :

- Variabel X1 adalah Profitabilitas
- Variabel X2 adalah Likuiditas & Efisiensi Aset
- Variabel X3 adalah Efisiensi Keuangan Bersih
- Variabel X4 adalah Struktur Modal & Laba Operasi
- Variabel X5 adalah Beban Bunga & Nilai Saham

- Variabel X6 adalah Efisiensi Modal & Pertumbuhan Aset
- Variabel X7 adalah Profitabilitas Saham
- Variabel X8 adalah Kesehatan Keuangan
- Variabel X9 adalah Leverage & Pertumbuhan Laba
- Variabel X10 adalah Posisi Ekonomi Makro

Interpretasi yang didapat sebagai berikut:

1. Profitabilitas (Koefisien: -0.24783058)  
Koefisien negatif menunjukkan bahwa semakin tinggi profitabilitas suatu perusahaan, semakin kecil kecenderungan perusahaan tersebut masuk ke dalam kelompok yang menjadi perusahaan bangkrut. Artinya, semakin besar laba dan pertumbuhan laba serta arus kas, semakin rendah kemungkinan perusahaan tergolong dalam kelompok resiko terkena bangkrut.
2. Likuiditas & Efisiensi Aset (Koefisien: 0.46350082)  
Koefisien positif dan cukup besar menunjukkan bahwa perusahaan dengan likuiditas tinggi dan efisiensi aset yang baik justru cenderung masuk ke kelompok berisiko. Ini bisa berarti

bahwa meskipun perusahaan terlihat mampu membayar kewajiban jangka pendek dan menggunakan asetnya dengan efisien, hal itu belum tentu cukup untuk menghindari masalah keuangan. Bisa jadi efisiensi tersebut hanya terlihat baik di permukaan, tapi ada masalah lain yang lebih besar di dalam.

3. Efisiensi Keuangan Bersih (Koefisien: 1.64381039)

Koefisien positif menunjukkan bahwa semakin tinggi efisiensi keuangan bersih perusahaan yakni kemampuan perusahaan menghasilkan keuntungan setelah memperhitungkan seluruh pendapatan dan pengeluaran—semakin besar pula kemungkinan perusahaan tergolong bangkrut. Ini mengindikasikan bahwa efisiensi keuangan bersih yang tinggi dalam konteks ini justru berkorelasi positif dengan risiko kebangkrutan. Kemungkinan besar hal ini terjadi karena efisiensi tersebut dicapai melalui pemangkasan biaya secara agresif, penundaan investasi penting, atau strategi keuangan berisiko lainnya yang dapat melemahkan kesehatan jangka panjang perusahaan.

4. Struktur Modal & Laba Operasi (Koefisien: 1.38395647)

Koefisien positif menunjukkan bahwa semakin tinggi struktur modal atau laba operasi perusahaan, semakin besar kecenderungan perusahaan untuk tergolong dalam kelompok bangkrut. Hal ini bisa disebabkan oleh ketergantungan pada pembiayaan berbasis utang (struktur modal tinggi) yang meningkatkan beban bunga dan risiko keuangan, atau oleh laba operasi yang tinggi namun tidak stabil. Ketidakstabilan tersebut dapat memperburuk posisi keuangan dalam jangka panjang dan meningkatkan risiko kebangkrutan, meskipun tampaknya menguntungkan dalam jangka pendek.

5. Beban Bunga & Nilai Saham (Koefisien: 0.61950317)

Koefisien positif menunjukkan bahwa semakin tinggi beban bunga atau nilai saham perusahaan, semakin besar kemungkinan perusahaan tergolong bangkrut. Beban bunga yang tinggi mencerminkan tekanan keuangan dari utang, sementara nilai saham yang tinggi namun tidak ditopang oleh kinerja keuangan yang sehat dapat menandakan overvaluasi dan risiko

koreksi pasar. Kombinasi keduanya dapat memperbesar risiko kebangkrutan.

6. Efisiensi Modal & Pertumbuhan Aset (Koefisien:  $-0.03397322$ )

Koefisien negatif menunjukkan bahwa semakin efisien penggunaan modal dan semakin tinggi pertumbuhan aset, semakin kecil kemungkinan perusahaan mengalami kebangkrutan. Ini mengindikasikan bahwa optimalisasi aset dan modal mendorong kinerja keuangan yang sehat, serta meningkatkan ketahanan perusahaan terhadap risiko kebangkrutan.

7. Profitabilitas Saham (Koefisien:  $0.53748570$ )

Koefisien positif menunjukkan bahwa semakin tinggi profitabilitas saham yang disertai fluktuasi pendapatan, semakin besar kemungkinan perusahaan mengalami kebangkrutan. Ini mencerminkan bahwa meskipun profitabilitas tampak tinggi, ketidakstabilan atau variasi minor dalam pendapatan dapat menciptakan ketidakpastian arus kas dan meningkatkan risiko keuangan jangka panjang.

8. Kesehatan Keuangan (Koefisien:  $-0.14237134$ )

Koefisien negatif menunjukkan bahwa semakin baik kesehatan keuangan perusahaan, semakin kecil kemungkinan perusahaan mengalami kebangkrutan. Sebaliknya, kondisi keuangan yang lemah meningkatkan risiko kegagalan usaha. Hal ini mencerminkan bahwa manajemen keuangan yang sehat berperan penting dalam menjaga stabilitas dan kelangsungan hidup perusahaan.

9. Leverage & Pertumbuhan Laba (Koefisien:  $-0.05241772$ )

Koefisien negatif menunjukkan bahwa penggunaan leverage yang disertai dengan pertumbuhan laba justru menurunkan kemungkinan perusahaan mengalami kebangkrutan. Artinya, jika utang dimanfaatkan secara produktif untuk mendorong pertumbuhan laba, maka risiko kebangkrutan dapat ditekan. Hal ini mencerminkan bahwa strategi pembiayaan dengan leverage yang terukur dan menghasilkan laba berkelanjutan dapat memperkuat ketahanan finansial perusahaan.

10. Posisi Ekonomi Makro (Koefisien:  $1.21090834$ )

Koefisien positif menunjukkan bahwa

perubahan pada faktor ekonomi makro berkorelasi positif dengan kemungkinan kebangkrutan. Artinya, ketika kondisi ekonomi makro memburuk—seperti inflasi tinggi, suku bunga naik, atau perlambatan pertumbuhan ekonomi—risiko kebangkrutan perusahaan meningkat. Ini menegaskan bahwa ketidakpastian dan tekanan ekonomi eksternal dapat berdampak signifikan terhadap keberlangsungan bisnis.

Secara keseluruhan, model ini menunjukkan bahwa risiko kebangkrutan perusahaan dipengaruhi oleh kombinasi faktor internal dan eksternal yang kompleks. Meskipun profitabilitas dan kesehatan keuangan yang baik dapat menurunkan risiko kebangkrutan, faktor seperti efisiensi keuangan bersih yang tinggi, struktur modal berat, beban bunga besar, serta kondisi ekonomi makro yang kurang mendukung justru meningkatkan risiko tersebut. Oleh karena itu, perusahaan perlu mengelola struktur modal dan beban utangnya secara hati-hati, menjaga stabilitas operasional, serta selalu memantau kondisi ekonomi eksternal agar dapat mempertahankan kelangsungan usaha dan mengurangi peluang gagal bayar. Pendekatan manajemen risiko yang komprehensif dan berimbang antara efisiensi internal dan

kewaspadaan terhadap kondisi makro sangat penting untuk menjaga stabilitas dan daya tahan perusahaan.

#### 4.3.2 Evaluasi dan Visualisasi

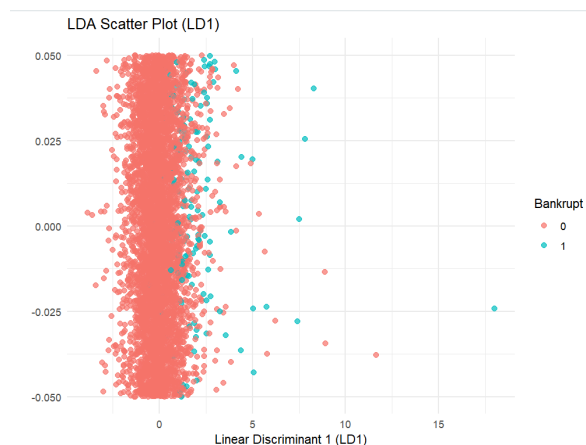
```
> conf_train
      Actual
Predicted 0    1
0  4475   96
1    85   41
> cat("Accuracy (Train):", sum(diag(conf_train)) / sum(conf_train),
"\n")
Accuracy (Train): 0.9614648
> conf_test
      Actual
Predicted 0    1
0  1868   35
1    31   17
> cat("Accuracy (Test):", sum(diag(conf_test)) / sum(conf_test), "\n")
Accuracy (Test): 0.9661712
```

Pada data train, model LDA menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan perusahaan yang tidak bangkrut, dengan 4475 data kelas 0 teridentifikasi secara benar. Namun, model masih kesulitan dalam mendeteksi kelas 1(bangkrut), dimana hanya 41 dari 137 kasus yang berhasil diprediksi dengan tepat, sedangkan 96 kasus lainnya salah diklasifikasikan sebagai tidak bangkrut. Akurasi train yang dihasilkan mencapai 96,15%, menunjukkan bahwa model memiliki performa tinggi, namun masih belum mencapai optimal untuk menangani kelas minoritas.

Sementara itu, pada data test, model juga menunjukkan pola serupa. Sebanyak 1.868 dari 1899 data kelas 0 berhasil diklasifikasikan dengan benar, tetapi

hanya 17 dari 52 data kelas 1 yang berhasil dikenali. Sebanyak 35 perusahaan bangkrut tidak terdeteksi (Falsa negatif), dan 31 data kelas 0 salah diprediksi sebagai bangkrut. Meski akurasi test mencapai 96,62%, model masih menunjukkan bias terhadap kelas mayoritas dan belum mampu mengklasifikasikan kelas minoritas secara efektif.

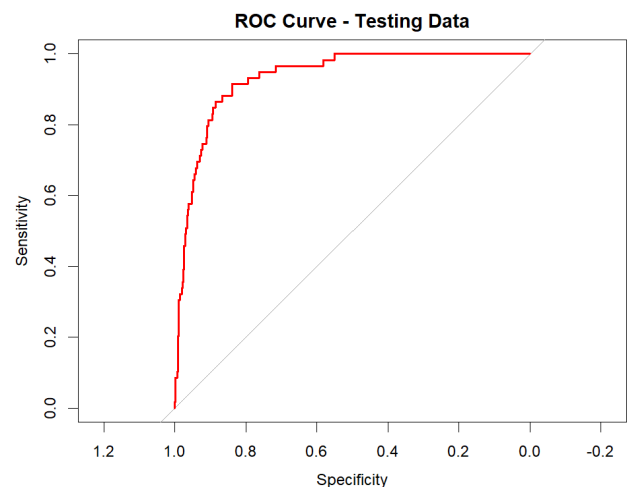
#### 4.3.3 Hasil dan Evaluasi Model



```
> conf_test
      Actual
Predicted 0    1
0      1910  43
1       21   16
> cat("Accuracy (Test):", sum(diag(conf_test)) / sum(conf_test),
"\n")
Accuracy (Test): 0.9678392
```

Dari total prediksi, model berhasil mengklasifikasikan 1.910 kasus non-bangkrut dan 16 kasus bangkrut dengan benar. Namun, terdapat kesalahan klasifikasi sebanyak 43 kasus non-bangkrut diprediksi sebagai bangkrut (false positive), dan 21 kasus

bangkrut diprediksi sebagai non-bangkrut (false negative). Akurasi model dihitung sebagai rasio jumlah prediksi yang benar terhadap total seluruh prediksi, yaitu sekitar 96.78%, yang menunjukkan performa klasifikasi yang sangat baik secara umum.



```
> cat("AUC (Test):", auc(roc_test), "\n")
AUC (Test): 0.9333093
```

Nilai AUC (Test) sebesar 0.933 menunjukkan bahwa model LDA Anda juga cukup baik dalam memprediksi status kebangkrutan pada data testing yang belum pernah dilihat model sebelumnya.

### 4.3 Regresi Logistik

#### 4.3.1 Uji Asumsi Non Multikolinearitas

```
> vif(model_vif)
Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7 Dim.8 Dim.9 Dim.10
1.095869 1.000059 1.003394 1.000392 1.000090 1.000126 1.000102 1.000727 1.008999 1.000274
Bangkrut
1.110032
```



Berdasarkan hasil output diperoleh nilai VIF Dim 1 hingga Dim 10 adalah <5 sehingga tidak terjadi multikolinieritas.

#### 4.3.2 Model Regresi Logistik

```
> n <- nrow(data_logistik)
> index_train <- sample(1:n, size = 0.7 * n)
> train_data <- data_logistik[index_train, ]
> test_data <- data_logistik[-index_train, ]
```

test_data	2046 obs. of 11 variables
train_data	4773 obs. of 11 variables

Dengan pembagian 70% data train dan 30% data test didapat jumlah data test adalah 2046 data dan data train berjumlah 4773.

```
> summary(model_train)

Call:
glm(formula = Bankrupt ~ ., family = binomial(), data = train_data)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.96610    0.18646  -26.634  <2e-16 ***
Dim.1       -0.73488    0.06090  -12.067  <2e-16 ***
Dim.2        0.12515    0.09186    1.362   0.1731
Dim.3        0.01854    0.24309    0.076   0.9392
Dim.4        0.35097    0.18440    1.903   0.0570 .
Dim.5        0.00336    0.23940    0.014   0.9888
Dim.6        0.38328    0.74487    0.515   0.6069
Dim.7        0.97929    0.53250    1.839   0.0659 .
Dim.8       -0.12326    0.16322   -0.755   0.4502
Dim.9       -0.68913    0.28885   -2.386   0.0170 *
Dim.10      -1.65664    1.79010   -0.925   0.3547
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1341.52  on 4794  degrees of freedom
Residual deviance: 900.64  on 4784  degrees of freedom
AIC: 922.64

Number of Fisher Scoring iterations: 8
```

Berdasarkan hasil output RStudio dapat dibentuk model regresi logistik, yaitu:

$$Y = -4.96610 - 0.73488X_1 + 0.12515X_2 + 0.01854X_3 + 0.35097X_4 + 0.00336X_5 + 0.38328X_6 + 0.97929X_7 - 0.12326X_8 - 0.68913X_9 - 1.65664X_{10}$$

1. Intercept (Konstanta : -4.96610)

Intercept sebesar -4.96610 dalam

model regresi logistik ini menunjukkan bahwa ketika seluruh variabel independen ( $X_1$  hingga  $X_{10}$ ) berada pada nilai nol, maka log-odds atau kecenderungan dasar perusahaan untuk mengalami kebangkrutan sangat rendah. Dengan kata lain, model mengasumsikan bahwa dalam keadaan netral (saat seluruh indikator risiko berada di titik nol), perusahaan secara umum tidak berada dalam posisi yang rentan terhadap kebangkrutan. Namun, nilai ini hanya menjadi titik awal, karena setiap variabel lain dalam model akan menambah atau mengurangi risiko tersebut tergantung pada arah dan besarnya koefisiennya.

2. Profitabilitas (Koefisien: -0.73488)

Koefisien negatif menunjukkan bahwa semakin tinggi profitabilitas perusahaan, semakin kecil kemungkinan perusahaan mengalami kebangkrutan. Profitabilitas yang baik mencerminkan kemampuan perusahaan menghasilkan laba yang cukup untuk menutupi biaya dan kewajiban, sehingga memperkuat posisi keuangan dan mengurangi risiko gagal bayar.

3. Likuiditas & Efisiensi Aset (Koefisien: 0.12515)

Koefisien positif menandakan bahwa

peningkatan likuiditas dan efisiensi aset sedikit meningkatkan kemungkinan kebangkrutan. Hal ini mungkin mengindikasikan bahwa perusahaan dengan likuiditas tinggi namun efisiensi aset yang kurang optimal berpotensi menghadapi risiko likuiditas jangka panjang atau pengelolaan aset yang kurang efektif.

4. Efisiensi Keuangan Bersih (Koefisien: 0.01854)

Koefisien positif kecil ini menunjukkan bahwa efisiensi keuangan bersih memiliki pengaruh ringan namun positif terhadap risiko kebangkrutan. Artinya, efisiensi keuangan yang tampak baik belum tentu menjamin kestabilan keuangan perusahaan secara keseluruhan.

5. Struktur Modal & Laba Operasi (Koefisien: 0.35097)

Koefisien positif mengindikasikan bahwa semakin besar struktur modal dan laba operasi, risiko kebangkrutan cenderung meningkat. Hal ini bisa terjadi jika perusahaan menggunakan utang yang berlebihan atau laba operasi yang fluktuatif, sehingga meningkatkan tekanan keuangan

6. Beban Bunga & Nilai Saham (Koefisien: 0.00336)

Koefisien positif sangat kecil ini menunjukkan bahwa beban bunga dan nilai saham memberikan dampak minimal namun tetap positif terhadap risiko kebangkrutan. Beban bunga yang tinggi dapat menekan arus kas, sementara nilai saham yang tinggi tidak selalu mencerminkan kondisi fundamental perusahaan.

7. Efisiensi Modal & Pertumbuhan Aset (Koefisien: 0.38328)

Koefisien positif ini menunjukkan bahwa efisiensi modal dan pertumbuhan aset yang lebih tinggi berhubungan dengan peningkatan risiko kebangkrutan. Mungkin perusahaan menggunakan modal secara agresif yang meningkatkan risiko keuangan jangka pendek.

8. Profitabilitas Saham (Koefisien: 0.97929)

Koefisien positif cukup besar ini menunjukkan bahwa profitabilitas saham yang tinggi justru berkorelasi dengan risiko kebangkrutan yang lebih besar, kemungkinan karena fluktuasi harga saham yang tajam atau ketidakstabilan pasar yang mempengaruhi kondisi keuangan perusahaan.

9. Kesehatan Keuangan (Koefisien: -0.12326)

Koefisien negatif ini menandakan bahwa kesehatan keuangan yang baik menurunkan risiko kebangkrutan. Kondisi keuangan yang sehat memperkuat posisi perusahaan dalam menghadapi tekanan bisnis dan risiko eksternal.

10. Leverage & Pertumbuhan Laba (Koefisien: -0.68913)

Koefisien negatif menunjukkan bahwa penggunaan leverage yang disertai dengan pertumbuhan laba menurunkan kemungkinan kebangkrutan. Hal ini mencerminkan strategi pembiayaan yang produktif dan pertumbuhan laba yang berkelanjutan meningkatkan daya tahan perusahaan.

11. Posisi Ekonomi Makro (Koefisien: -1.65664)

Koefisien negatif signifikan ini menunjukkan bahwa kondisi ekonomi makro yang mendukung secara signifikan mengurangi risiko kebangkrutan perusahaan. Faktor eksternal seperti inflasi rendah, suku bunga stabil, dan pertumbuhan ekonomi positif membantu perusahaan bertahan dalam jangka panjang.

Model regresi logistik ini mengimplikasikan bahwa risiko kebangkrutan perusahaan dipengaruhi oleh kombinasi faktor internal dan eksternal, dengan pengaruh terbesar berasal dari profitabilitas saham dan kondisi ekonomi makro. Faktor-faktor seperti profitabilitas, kesehatan keuangan, dan leverage disertai pertumbuhan laba berperan menurunkan risiko kebangkrutan, sementara struktur modal, efisiensi modal, dan fluktuasi saham cenderung meningkatkan risiko. Intercept yang negatif menunjukkan bahwa secara umum, tanpa tekanan dari variabel-variabel tersebut, risiko kebangkrutan berada pada tingkat yang rendah. Maka, perusahaan perlu menjaga stabilitas keuangan internal sambil mengantisipasi dinamika ekonomi eksternal untuk meminimalkan kemungkinan bangkrut.

### 4.3.3 Uji Signifikansi Keseluruhan Model

```
> pr2(model_train)
fitting null model for pseudo-r2
      11h      11hNu11      G2      McFadden      r2ML
-450.32248379 -670.76248278 440.87999798 0.32864092 0.08784539
      r2CU
0.35995307
> qchisq(0.95,9)
[1] 16.91898
```

Berdasarkan hasil output RStudio diperoleh G2 sebesar 440.87999798

dengan nilai chi square tabel sebesar 16.91898. Dikarenakan nilai G2 lebih besar dibandingkan dengan nilai chi square maka dapat diputuskan bahwa H0 ditolak. Sehingga dapat disimpulkan bahwa model signifikan atau minimal terdapat satu variabel yang berpengaruh terhadap terjadinya kebangkrutan pada perusahaan.

#### 4.3.4 R square

Rsq	0.328642137277118
-----	-------------------

R square dihitung dengan membagi nilai Residual deviance dan Null deviance. Berdasarkan hasil output RStudio diperoleh nilai R square adalah 0.32864. Dari Nilai R square dapat diketahui bahwa variabel independen (Dim1-Dim10) dapat menjelaskan 32.86% potensi kebangkrutan suatu perusahaan.

#### 4.3.5 Odds Ratio

```
> beta<-(coef(model_train))
> beta
      (Intercept)      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
-4.966094794 -0.734883216  0.125147068  0.018543436  0.350969825  0.003360367
      Dim.6      Dim.7      Dim.8      Dim.9      Dim.10
 0.383280892  0.979286112 -0.123256147 -0.689128090 -1.656642044
> OR_beta<-exp(beta)
> OR_beta
      (Intercept)      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
 0.006970315  0.479561461  1.133315115  1.018716433  1.420444463  1.003366019
      Dim.6      Dim.7      Dim.8      Dim.9      Dim.10
 1.467090066  2.662554797  0.884037190  0.502013589  0.190778532
> cbind(beta,OR_beta)
      beta      OR_beta
(Intercept) -4.966094794 0.006970315
Dim.1       -0.734883216 0.479561461
Dim.2       0.125147068 1.133315115
Dim.3       0.018543436 1.018716433
Dim.4       0.350969825 1.420444463
Dim.5       0.003360367 1.003366019
Dim.6       0.383280892 1.467090066
Dim.7       0.979286112 2.662554797
Dim.8       -0.123256147 0.884037190
Dim.9       -0.689128090 0.502013589
Dim.10      -1.656642044 0.190778532
```

Interpretasi dari nilai odds ratio sebagai berikut :

##### 1. Profitabilitas (OR: 0.4796)

Jika profitabilitas meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan menurun menjadi 0.480 kali lipat. Artinya, semakin tinggi profitabilitas, semakin kecil kemungkinan perusahaan bangkrut karena keuntungan yang stabil memberikan bantalan keuangan yang kuat.

##### 2. Likuiditas & Efisiensi Aset (OR: 1.1333)

Jika likuiditas dan efisiensi aset meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan meningkat 1.133 kali lipat. Mengindikasikan bahwa meskipun efisiensi aset dan likuiditas meningkat, jika tidak disertai

pengelolaan yang sehat, bisa memperbesar risiko.

3. Efisiensi Keuangan Bersih (OR: 1.0187)

Jika efisiensi keuangan bersih meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan meningkat 1.018 kali lipat. Kenaikan ini relatif kecil, menandakan bahwa faktor ini memiliki pengaruh yang minim, namun tetap perlu diawasi sebagai bagian dari pengelolaan biaya dan pendapatan bersih.

4. Struktur Modal & Laba Operasi (OR: 1.4204)

Jika struktur modal dan laba operasi meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan meningkat 1.420 kali lipat. Ini mencerminkan bahwa perusahaan dengan struktur modal yang kompleks atau laba operasi yang tidak stabil cenderung lebih berisiko mengalami tekanan finansial.

5. Beban Bunga & Nilai Saham (OR: 1.0034)

Jika beban bunga dan nilai saham meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan meningkat 1.003 kali

lipat. Efek ini sangat kecil, namun tetap menunjukkan bahwa peningkatan beban bunga bisa menambah tekanan likuiditas.

6. Efisiensi Modal & Pertumbuhan Aset (OR: 1.4671)

Jika efisiensi modal dan pertumbuhan aset meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan meningkat 1.467 kali lipat. Hal ini bisa terjadi jika ekspansi aset dilakukan terlalu agresif tanpa memperhitungkan risiko keuangan yang menyertainya.

7. Profitabilitas Saham (OR: 2.6626)

Jika profitabilitas saham meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan meningkat 2.662 kali lipat. Ini mengindikasikan bahwa fluktuasi laba per saham yang tinggi bisa menjadi tanda ketidakstabilan kinerja jangka pendek yang berdampak buruk bagi kelangsungan usaha.

8. Kesehatan Keuangan (OR: 0.8840)

Jika kesehatan keuangan meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan menurun sebesar 0.884 kali lipat. Ini berarti kondisi internal keuangan yang

baik tetap menjadi faktor pelindung terhadap risiko kebangkrutan.

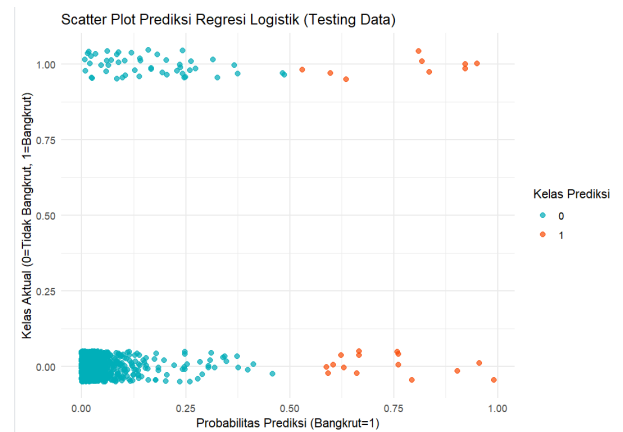
#### 9. Leverage & Pertumbuhan Laba (OR: 0.5020)

Jika leverage dan pertumbuhan laba meningkat 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan menurun sebesar 0.502 kali lipat. Penggunaan utang yang produktif dan diiringi dengan pertumbuhan laba menjadi strategi yang mendukung kelangsungan bisnis.

#### 10. Posisi Ekonomi Makro (OR: 0.1908)

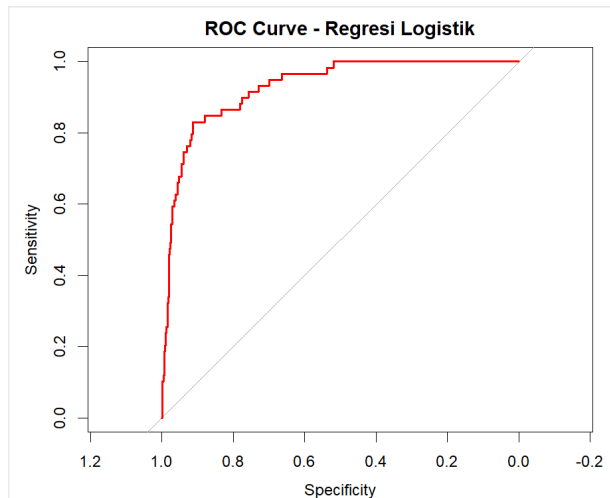
Jika kondisi ekonomi makro membaik 1 satuan, maka kecenderungan perusahaan mengalami kebangkrutan menurun sebesar 0.190 kali lipat. Ini menunjukkan bahwa stabilitas ekonomi eksternal merupakan penentu penting dalam menjaga kelangsungan hidup perusahaan di tengah ketidakpastian pasar.

### 4.3.6 Hasil dan Evaluasi



```
> table(Predicted = pred_class, Actual = test_data$Bankrupt)
      Actual
Predicted 0    1
      0 1916  50
      1   15   9
> # Akurasi
> accuracy <- mean(pred_class == as.numeric(test_data$Bankrupt) - 1)
> print(paste("Akurasi data test:", round(accuracy, 3)))
[1] "Akurasi data test: 0.967"
```

Dari total prediksi, model berhasil mengklasifikasikan 1.916 kasus non-bangkrut dan 15 kasus bangkrut dengan benar. Namun, terdapat kesalahan klasifikasi sebanyak 50 kasus non-bangkrut diprediksi sebagai bangkrut (false positive), dan 9 kasus bangkrut diprediksi sebagai non-bangkrut (false negative). Akurasi model dihitung sebagai rasio jumlah prediksi yang benar terhadap total seluruh prediksi, yaitu sekitar 96.7%, yang menunjukkan performa klasifikasi yang sangat baik secara umum.



```
> cat("AUC (Test):", auc(roc_test_logit), "\n")
AUC (Test): 0.9286661
```

Nilai AUC (Test) sebesar 0.929 menunjukkan bahwa model Regresi Logistik juga cukup baik dalam memprediksi status kebangkrutan pada data testing yang belum pernah dilihat model sebelumnya.

#### 4.4 Perbandingan Hasil Analisis Diskriminan Linear dan Regresi Logistik

Analisis perbandingan dilakukan untuk mengevaluasi keefektifan dua metode klasifikasi utama, yaitu Linear Discriminant Analysis (LDA) dan Regresi Logistik, dalam memprediksi kebangkrutan perusahaan berdasarkan sepuluh komponen utama hasil reduksi dimensi menggunakan PCA.

##### 4.4.1 Akurasi Klasifikasi

Berdasarkan hasil evaluasi pada data testing, kedua model menunjukkan performa akurasi yang tinggi. Model LDA mencatatkan akurasi sebesar 96.78%. Sementara itu, model Regresi Logistik mencatatkan akurasi sebesar 96.7% pada data testing. Hal ini menunjukkan bahwa secara umum, kedua metode memiliki kemampuan yang sangat baik dalam mengklasifikasikan perusahaan yang tidak bangkrut (kelas mayoritas).

##### 4.4.2 Kemampuan Mendeteksi Kelas Minoritas (Sensitivitas)

```
> cat("Sensitivitas:", round(sensitivitas_logit, 4), "\n")
Sensitivitas: 0.2414
> cat("Spesifisitas:", round(spesifisitas_logit, 4), "\n")
Spesifisitas: 0.996
```

##### Sensitivitas dan Spesifisitas Regresi Logistik

```
> sensitivitas_lda
[1] 0.2711864
> spesifisitas_lda
[1] 0.9891248
```

##### Sensitivitas dan Spesifisitas LDA

Berdasarkan hasil perbandingan antara regresi logistik dan analisis diskriminan linear (LDA) pada metrik sensitivitas dan spesifisitas, terlihat bahwa model regresi logistik memiliki sensitivitas sebesar 0.2414 dan spesifisitas sebesar

0.996, sedangkan model LDA memiliki sensitivitas sebesar 0.2712 dan spesifisitas sebesar 0.9891. Artinya, model LDA sedikit lebih baik dalam mendeteksi kasus bangkrut (positif) dibanding regresi logistik, meskipun model logistik sedikit lebih unggul dalam mengidentifikasi kasus tidak bangkrut (negatif). Dengan kata lain, LDA lebih sensitif terhadap kasus-kasus kebangkrutan, sedangkan regresi logistik lebih akurat dalam memastikan perusahaan yang tidak bangkrut.

#### 4.4.3 Interpretabilitas Model

Dari sisi interpretabilitas, LDA memberikan fungsi diskriminan linear yang secara eksplisit menunjukkan kontribusi dari setiap komponen utama terhadap klasifikasi. Hal ini memungkinkan peneliti untuk mengidentifikasi dimensi-dimensi yang paling berpengaruh dalam membedakan kelompok. Sebaliknya, Regresi Logistik menyediakan informasi Odds Ratio (OR), yang dapat digunakan untuk menilai dampak relatif dari setiap variabel terhadap kemungkinan kebangkrutan.

#### 4.4.4 Evaluasi Visualisasi dan ROC Curve

Berdasarkan perbandingan nilai AUC pada data uji (testing), model Linear Discriminant Analysis (LDA) memperoleh nilai AUC sebesar 0.933, sedangkan model Regresi Logistik menghasilkan nilai AUC sebesar 0.929. Kedua nilai tersebut berada pada kisaran yang sangat baik (mendekati 1), yang menunjukkan bahwa kedua model sama-sama memiliki kemampuan yang tinggi dalam membedakan antara perusahaan yang bangkrut dan tidak bangkrut. Namun, model LDA memiliki sedikit keunggulan dalam hal akurasi klasifikasi secara keseluruhan, sebagaimana tercermin dari nilai AUC-nya yang lebih tinggi, meskipun selisihnya sangat tipis.

**Table 3.1 Perbedaan model linear diskriminan analisis dan regresi logistik**

Kriteria	LDA	Regresi Logistik
a		
Akurasi (Test)	96.78%	96.7%



Sensitivitas	27.11%	24.14%
Spesifitas	98.91%	99.6%
AUC (Test)	0.933 (Cukup Baik)	0.929 (Cukup Baik)
Deteksi Bangkrut(1)	Lebih baik	Kurang optimal

## 9. KESIMPULAN

Model LDA dan Regresi Logistik sama-sama menunjukkan akurasi tinggi dalam klasifikasi status kebangkrutan, masing-masing sekitar 96.78% dan 96.7%. Namun, LDA memiliki sensitivitas lebih tinggi (27.11%), sehingga lebih baik dalam mendeteksi perusahaan yang benar-benar bangkrut. Di sisi lain, regresi logistik memiliki spesifisitas lebih tinggi (99.6%), lebih akurat dalam mengidentifikasi perusahaan yang tidak bangkrut.

Secara umum, LDA cocok digunakan saat fokus utama adalah mendeteksi risiko kebangkrutan, sedangkan regresi logistik lebih sesuai jika ingin meminimalkan kesalahan klasifikasi pada perusahaan yang sehat,

sekaligus memberikan interpretasi risiko melalui odds ratio.

## 10. SARAN

Bagi peneliti selanjutnya penting untuk mengembangkan penelitian ini dengan menggunakan

1. Mengembangkan pendekatan dengan analisis diskriminan lain, seperti Analisis Diskriminan Kuadratik (QDA) atau Canonical Discriminant Analysis (CDA), khususnya bila asumsi kesamaan kovarian antar kelompok tidak terpenuhi, agar fungsi klasifikasi lebih sesuai dengan karakteristik data.

2. Mengintegrasikan hasil PCA dengan analisis faktor eksploratori (EFA) untuk mengidentifikasi struktur laten dan kelompok indikator keuangan yang membentuk dimensi penting, sehingga hasil klasifikasi memiliki dasar konseptual yang lebih kuat.

3. Melakukan pengujian asumsi multivariat secara menyeluruh, seperti normalitas multivariat, homogenitas matriks kovarian, dan multikolinearitas antar prediktor, untuk memastikan

bahwa syarat-syarat penggunaan LDA atau regresi logistik dalam konteks multivariat benar-benar terpenuhi.

4. Memanfaatkan metode rotasi komponen (misalnya rotasi varimax) setelah PCA, agar setiap dimensi yang terbentuk lebih mudah diinterpretasikan dan lebih representatif terhadap konsep-konsep keuangan seperti profitabilitas, efisiensi, atau likuiditas.

6. Melakukan validasi silang (cross-validation) dalam kerangka multivariat untuk menghindari overfitting, serta menguji stabilitas fungsi diskriminan dengan membandingkan hasil klasifikasi pada data training dan testing secara seimbang.

7. Mengkaji ulang asumsi proporsi awal (prior probabilities) dalam analisis diskriminan, terutama bila terjadi ketimpangan kelas, untuk meningkatkan akurasi klasifikasi dan relevansi hasil model terhadap kondisi data sesungguhnya.

8. Membandingkan hasil analisis diskriminan dengan pendekatan multivariat lainnya, seperti MANOVA atau analisis klaster, untuk memahami

struktur dan pola kelompok data secara lebih komprehensif dan mendalam dari berbagai sudut pandang statistika klasik.

## REFERENSI

[1] Huang, Y., Yang, H., Xu, J., & Wang, J. (2020). Performance metrics for class-imbalanced data classification: A review. *IEEE Access*, 8, 70027–70049.

[2] Fawcett, T. (2020). *Performance evaluation methods for classification*. In *Pattern Recognition and Machine Learning* (pp. 167–188). Springer.

[3] Kitbumrungrat, Krieng 2012. Comparison Logistic Regression and Discriminant Analysis in classification groups for Breast Cancer [Online]. Termuat di: [http://paper.ijcsns.org/07\\_book/201205/20120518.pdf](http://paper.ijcsns.org/07_book/201205/20120518.pdf). [Diakses : 9 maret 2013]

[4] Nattino, G., Pennell, M. L., & Lemeshow, S. (2020). Evaluating the goodness-of-fit of logistic regression models in large samples: A modified Hosmer–Lemeshow test. *Statistics in*

*Medicine*, 39(3), 409–421.  
<https://doi.org/10.1002/sim.8404>

[5] Norton, E. C. (2018). *Odds Ratios—Current Best Practice and Use* [PDF]. *JAMA*.

[6] Talha Mahboob Alam, Kamran Shaukat, Mubbashar Mushtaq, Yasir Ali, Matloob Khushi, Suhuai Luo, Abdul Wahab, Corporate Bankruptcy Prediction: An Approach Towards Better Corporate World, *The Computer Journal*, Volume 64, Issue 11, November 2021, Pages 1731–1746, <https://doi.org/10.1093/comjnl/bxaa056>

[7] Li, L., & Faff, R. (2019). Predicting corporate bankruptcy: What matters? *International Review of Economics & Finance*, 62, 1-19.

[8] Jaffari, A. A., & Ghafoor, Z. (2017). Predicting corporate bankruptcy in Pakistan a comparative study of multiple discriminant analysis (MDA) and logistic regression. *Research Journal of Finance and Accounting*, 8(3), 81-100.

[10] Jones, S. Corporate bankruptcy prediction: a high dimensional analysis. *Rev Account Stud* 22, 1366–1422 (2017).

<https://doi.org/10.1007/s11142-017-9407-1>

[11] Nyitrai, T., & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*, 67, 34-42. doi:10.1016/j.seps.2018.08.004.

[12] Jones, S., Johnstone, D. and Wilson, R. (2017), Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. *Journal of Business Finance & Accounting*, 44: 3-34. <https://doi.org/10.1111/jbfa.12218>

[13] Le, T., Lee, M. Y., Park, J. R., & Baik, S. W. (2018). Oversampling Techniques for Bankruptcy Prediction: Novel Features from a Transaction Dataset. *Symmetry*, 10(4), 79. <https://doi.org/10.3390/sym10040079>

[14] Zhang, J. (2024). Impact of an improved random forest-based financial management model on the effectiveness of corporate sustainability decisions. *Systems and Soft Computing*, 6. doi:10.1016/j.sasc.2024.200102.

[15] Affes, Z., Hentati-Kaffel, R. Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis. *Comput Econ* 54, 199–244 (2019). <https://doi.org/10.1007/s10614-017-9698-0>

[16] Hayati, E., Novitasari, D. A., & Rosdiyati, R. (2018). Analisis Diskriminan Dan Multivariate Adaptive Regression Spline (Mars) Untuk Memprediksi Financial Distress Pada Perusahaan Manufaktur Di BEI. *Media Mahardhika*, 17(1), 16-24.

[17] Wulandari, L., & Kurniawan, A. (2018). *Penerapan Regresi Logistik dalam Penentuan Faktor Risiko Hipertensi*. *Jurnal Statistika Terapan*, 5(1), 57–66.

[18] Susanti, R. (2018). *Analisis Regresi Logistik pada Data Risiko Kejadian Diabetes*. *Jurnal Matematika dan Sains*, 9(1), 10–18.

[19] Lee, S. (2025, March 13). *Mastering pseudo R-squared metrics in logistic regression models*. *Number Analytics Blog*.