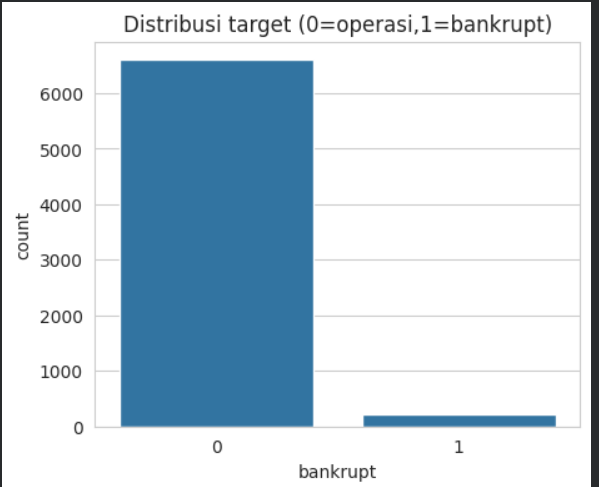
**2331020\_Ripto**

**1. Uraian Masalah dan Urgensi Predictive Analytics**

Bank Sentral Taiwan memiliki mandat untuk menjaga stabilitas sistem keuangan dengan memastikan seluruh bank berada dalam kondisi sehat. Risiko kebangkrutan bank dapat mengganggu stabilitas nasional karena menyebabkan bank run, kerugian deposan, dan efek penularan antarbank. Dataset yang tersedia memuat indikator kesehatan bank beserta label bangkrut atau tidak. Predictive analytics diperlukan untuk membangun sistem peringatan dini sehingga Bank Sentral dapat melakukan intervensi lebih cepat. Proyek ini berkontribusi melalui identifikasi faktor risiko utama serta penyediaan model prediksi yang dapat digunakan dalam pengawasan berkelanjutan.

**2. Data Understanding dan Exploratory Data Analysis**

Target variabel adalah ‘bankrupt’ yang bernilai 0 untuk bank yang masih beroperasi dan 1 untuk bank yang bangkrut. Hasil EDA menunjukkan distribusi kelas sangat tidak seimbang dengan sekitar 96.77% kelas 0 dan 3.23% kelas 1. Visualisasi seperti countplot, korelasi absolut, histogram, dan boxplot menunjukkan bahwa rasio profitabilitas, leverage, dan struktur pendanaan memiliki hubungan kuat dengan status kebangkrutan. Beberapa fitur dengan korelasi tinggi meliputi net\_income\_to\_total\_assets, ROA, net\_worth/assets, dan rasio profitabilitas lain.  
  
A screenshot of a computer

AI-generated content may be incorrect.A graph with a bar chart

AI-generated content may be incorrect.A graph with a bar chart

AI-generated content may be incorrect.A graph with numbers and a bar chart

AI-generated content may be incorrect.A graph with a number of blue and black lines

AI-generated content may be incorrect.A graph with a number of lines and numbers

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

**3. Data Preparation**

Pengecekan missing value menunjukkan tidak ada nilai hilang signifikan sehingga tidak ada kolom yang dihapus. Data sangat imbalanced sehingga diterapkan teknik SMOTE pada data latih. Pembagian data menggunakan train-test split dengan stratifikasi. Untuk model ANN dilakukan standardisasi fitur menggunakan StandardScaler agar pelatihan stabil. Secara keseluruhan, tahap data preparation mencakup pembersihan data, penyeimbangan kelas, pembagian data, dan transformasi fitur.

Pemeriksaan keseimbangan (balance):

- Periksa proporsi kelas target. Jika proporsi kelas minoritas jauh lebih kecil (mis. <10% atau <20%), dataset dianggap tidak seimbang.

Penanganan data tidak seimbang:

- Teknik sampling: SMOTE (oversampling sintetis pada data pelatihan), undersampling, atau kombinasi (SMOTE + Tomek).

- Alternatif: gunakan parameter class\_weight='balanced' pada model (RandomForest, DecisionTree) atau threshold tuning pada output probabilitas.

Tahapan cleaning & transformation yang diterapkan:

- Hapus/rename kolom, bersihkan nama kolom (lowercase, ganti spasi dengan underscore).

- Hapus kolom dengan >50% missing (opsional, tergantung konteks).

- Imputasi nilai hilang numerik menggunakan median (lebih robust terhadap outlier).

- Standarisasi/Scaling (StandardScaler) untuk ANN; Random Forest tidak membutuhkan scaling tetapi pipeline konsisten menggunakan scaler agar mudah integrasi model.

- Encoding untuk fitur kategorikal (One-Hot Encoder) jika ada.

**4. Modelling**

Tiga model digunakan: Decision Tree, Random Forest dengan SMOTE, dan Artificial Neural Network. Decision Tree menjadi baseline sederhana. Random Forest digunakan sebagai model utama dan memberi hasil feature importance. ANN dibuat dengan feedforward neural network menggunakan lapisan Dense, Dropout, dan aktivasi sigmoid. Random Forest dipilih karena robust dan mampu menangani fitur numerik dalam jumlah besar.

Implementasi singkat:

- Split data: train/test (contoh 80/20) dengan stratify pada target.

- Untuk Random Forest & Decision Tree: gunakan pipeline dengan SMOTE + scaler (scaler optional untuk tree). Gunakan class\_weight='balanced' jika mau alternatif SMOTE.

- Untuk ANN: lakukan SMOTE hanya pada data pelatihan, lalu scaling (StandardScaler), kemudian architecture sederhana 2 layer dense (128,64) dengan dropout, output sigmoid, loss=binary\_crossentropy.

A screenshot of a computer

AI-generated content may be incorrect.

**5. Evaluation**

Evaluasi dilakukan menggunakan akurasi, precision, recall, F1-score, confusion matrix, dan ROC AUC. Random Forest memberikan kinerja terbaik dengan akurasi sekitar 96%, recall kelas bangkrut 0.61, dan ROC AUC 0.94. Decision Tree lebih lemah dengan akurasi sekitar 89% dan precision kelas bangkrut rendah. ANN memiliki akurasi tinggi namun recall kelas bangkrut rendah (0.39). Dengan demikian, Random Forest merupakan model paling sesuai untuk prediksi kebangkrutan.

A screenshot of a graph

AI-generated content may be incorrect. A screenshot of a graph

AI-generated content may be incorrect. A screenshot of a graph

AI-generated content may be incorrect.

**6. Variabel Paling Mempengaruhi Kebangkrutan Bank**

Berdasarkan feature importance, lima variabel paling berpengaruh adalah:

1) net\_income\_to\_total\_assets

2) borrowing\_dependency

3) persistent\_eps\_in\_the\_last\_four\_seasons

4) after-tax\_net\_interest\_rate

5) continuous\_interest\_rate\_(after\_tax)

Kelima indikator ini terkait erat dengan profitabilitas, ketergantungan pendanaan, stabilitas pendapatan, dan struktur pendanaan bank, yang secara teori memang menjadi faktor penting risiko kebangkrutan.