Dataset Characteristics

The dataset used in this study consists of 6,819 firm-level observations and 96 variables, including 95 financial ratio indicators and one binary target variable (Bankrupt?). The dataset exhibits excellent data quality, with no missing values detected across all variables. The target variable is binary, where a value of 1 indicates a bankrupt company and 0 indicates a non-bankrupt company.

Class Imbalance Analysis

Exploratory analysis reveals a severe class imbalance in the target variable. Approximately 3% of the companies are classified as bankrupt, resulting in an imbalance ratio of roughly 1:30 (bankrupt to non-bankrupt firms).

This imbalance has important implications for model development and evaluation:

- Overall accuracy is not a reliable performance metric
- Models may become biased toward the majority (non-bankrupt) class
- Special techniques such as Synthetic Minority Oversampling Technique (SMOTE) or class-weighted algorithms are required
- Recall (Sensitivity) must be prioritized to minimize false negatives

Business implication: Failing to identify a bankrupt company represents a significantly higher cost than falsely flagging a healthy firm.

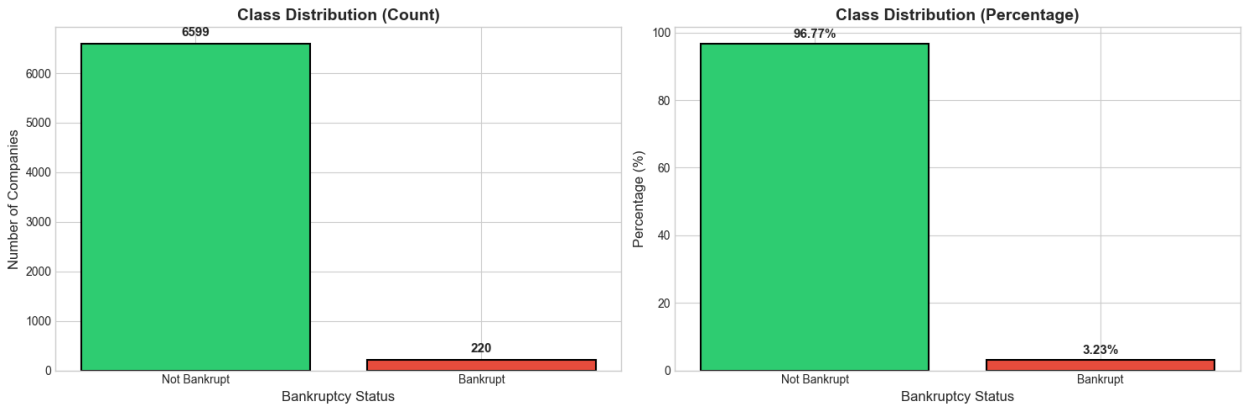


Figure 1: *Distribution of Bankrupt and Non-Bankrupt Companies (Bar chart showing severe class imbalance)*

Feature Grouping and Financial Interpretation

The 95 financial indicators were grouped into five major financial dimensions based on financial theory and prior bankruptcy literature:

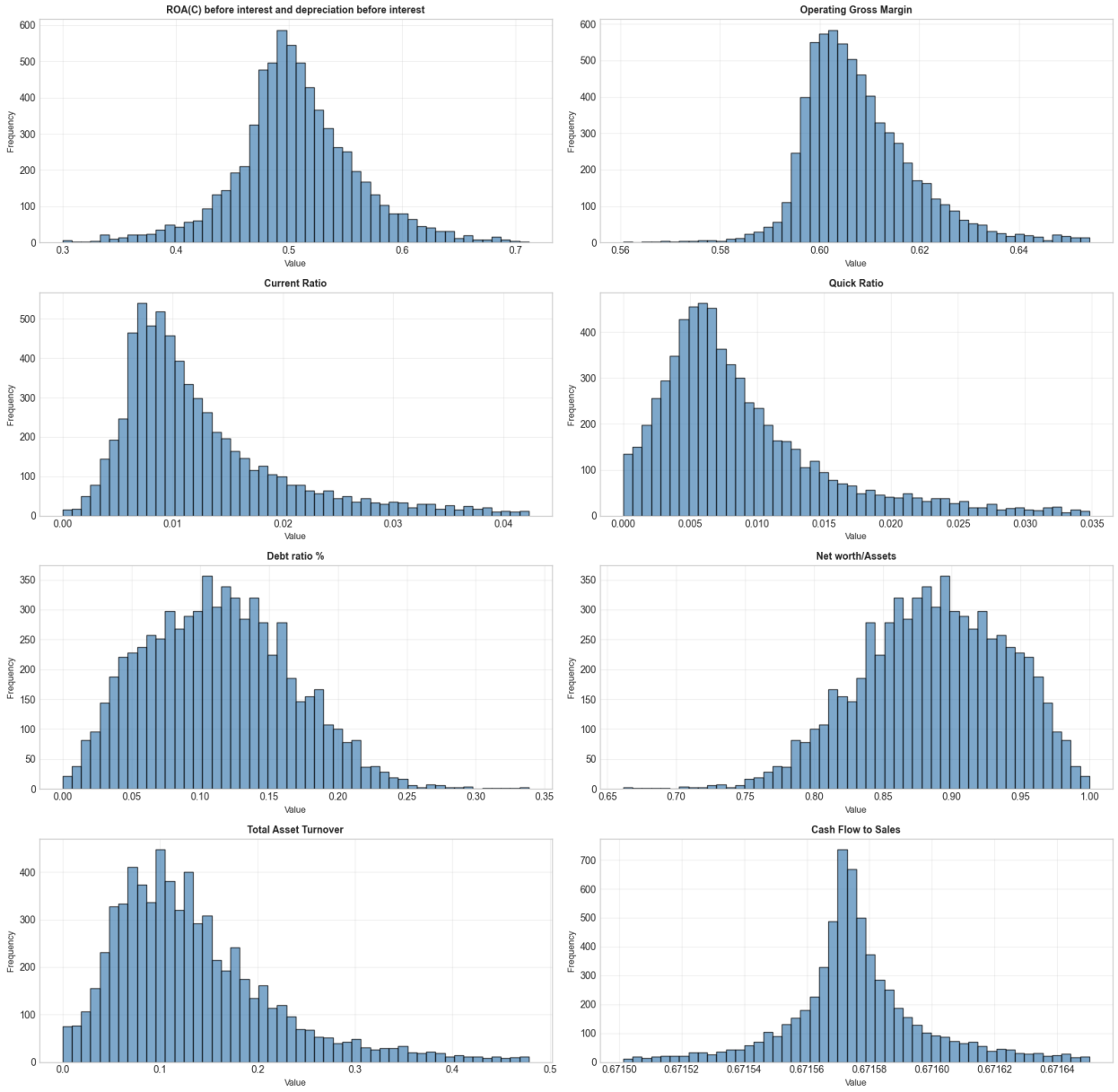
1. Profitability Ratios – e.g., ROA, ROE, operating profit margins
2. Liquidity Ratios – e.g., current ratio, quick ratio, cash ratios
3. Solvency Ratios – e.g., debt ratio, equity ratios
4. Efficiency Ratios – e.g., asset turnover, inventory turnover
5. Cash Flow Indicators – e.g., operating cash flow ratios

This categorization enables structured analysis and supports the development of an interpretable predictive framework.

Discriminative Power of Features

Several financial ratios demonstrate strong discriminative ability between bankrupt and non-bankrupt firms. Visual inspection using box plots shows clear separation in median values for key profitability, liquidity, and leverage ratios. In many cases, bankrupt firms exhibit lower profitability, weaker liquidity positions, and higher leverage.

Distribution of Key Financial Ratios



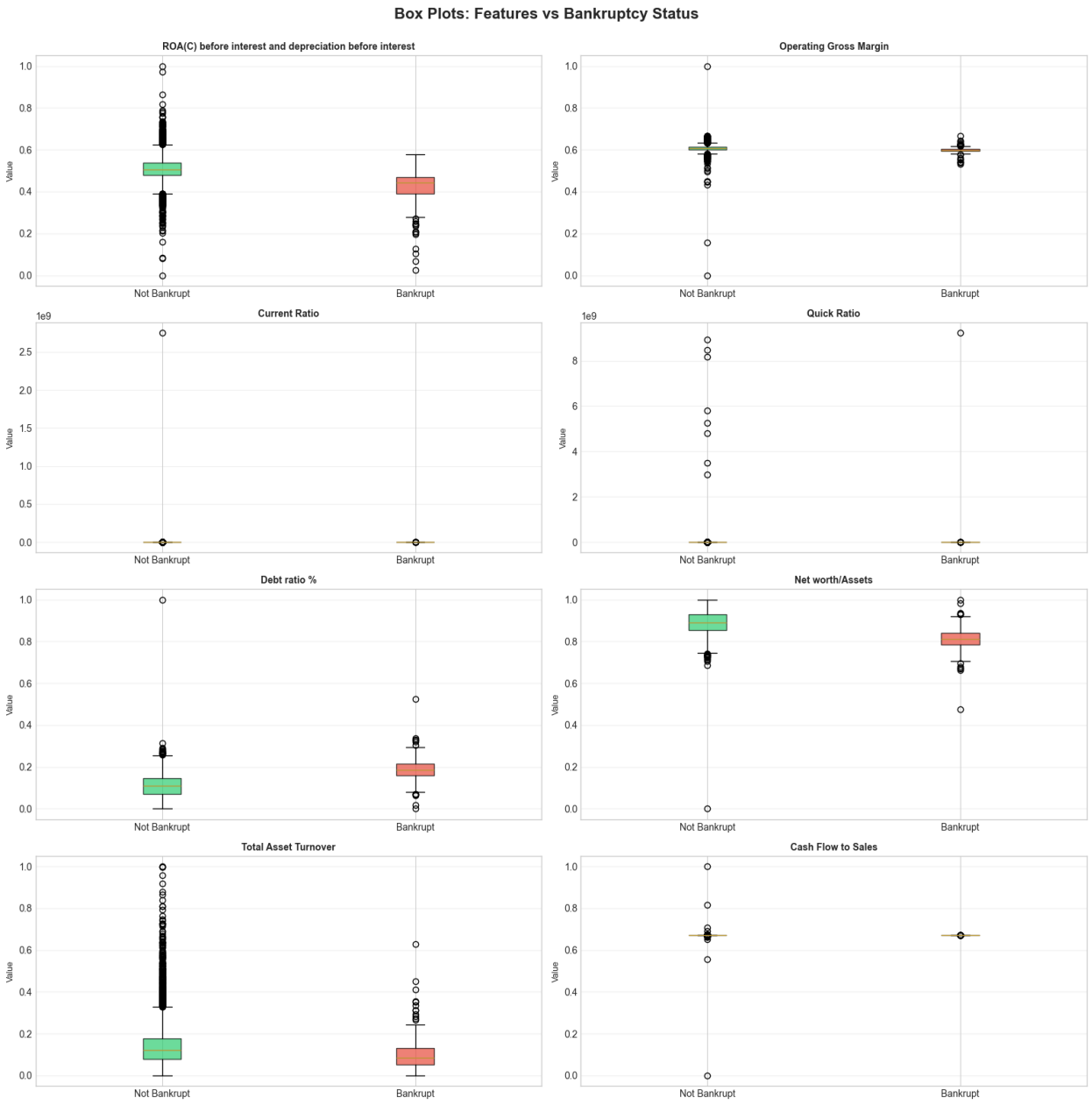
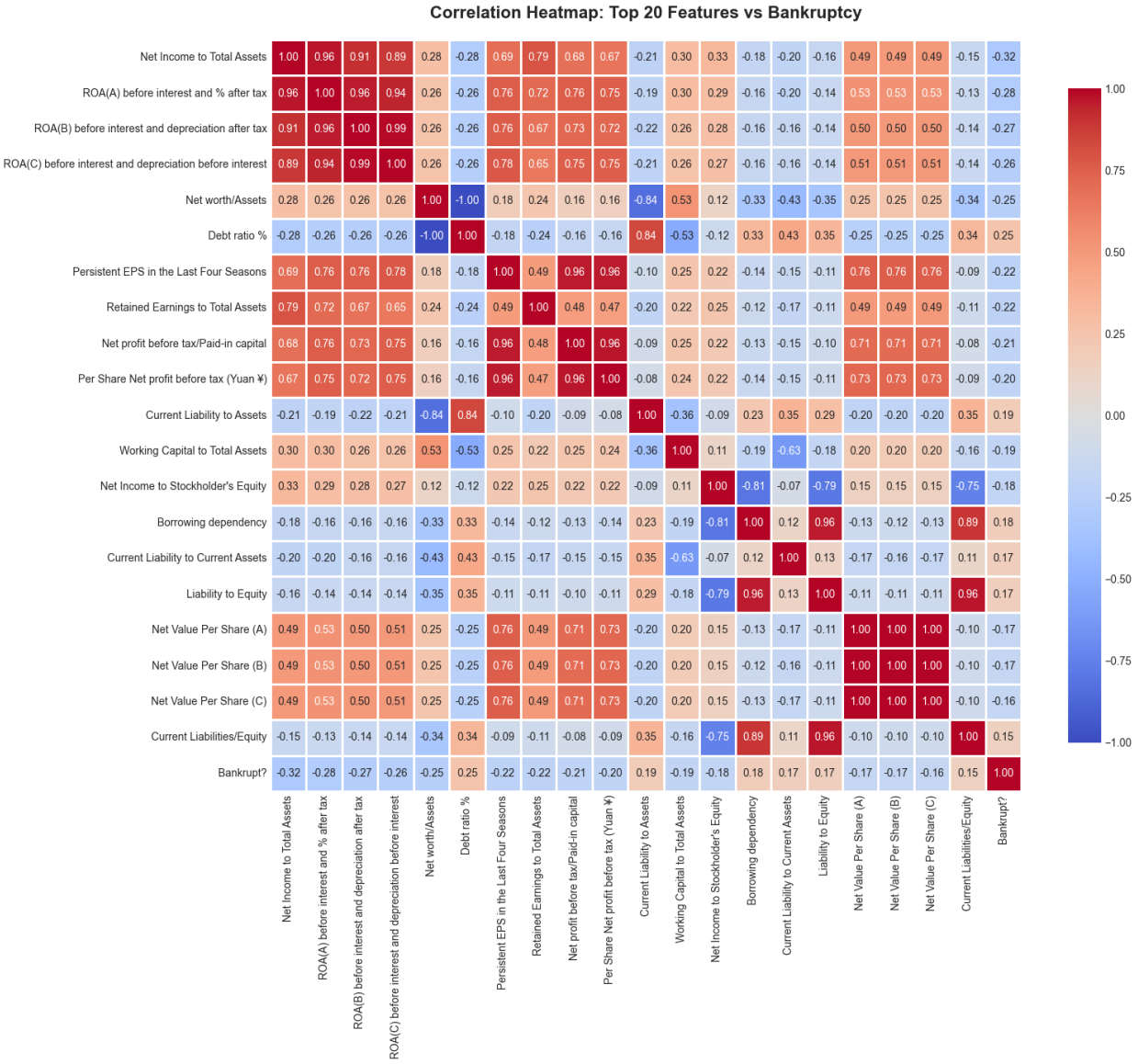
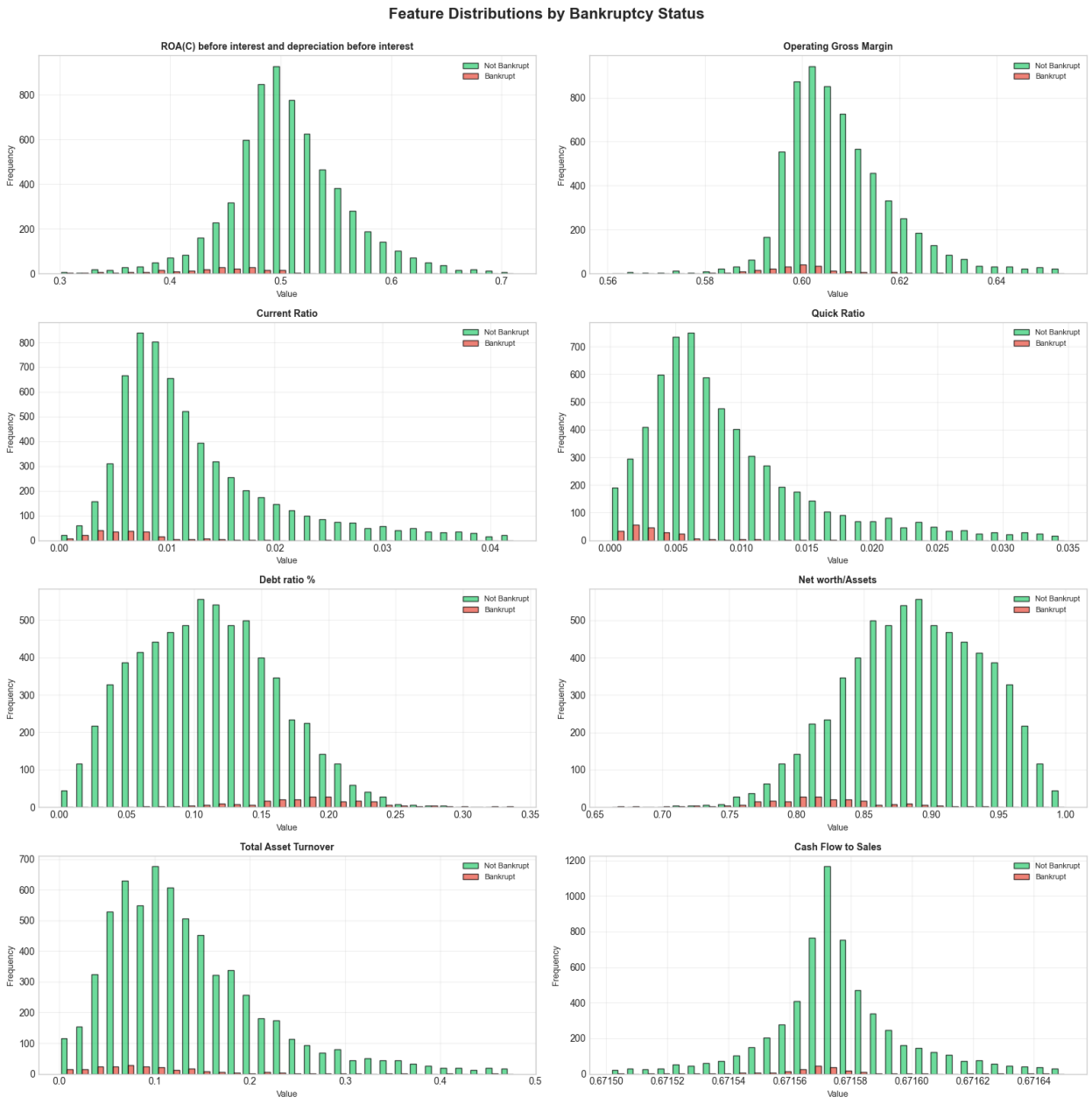


Figure 3: *Box Plot Comparison of Liquidity and Leverage Ratios (Example: Current Ratio, Debt Ratio %)*



Top 10 Features Correlated with Bankruptcy:

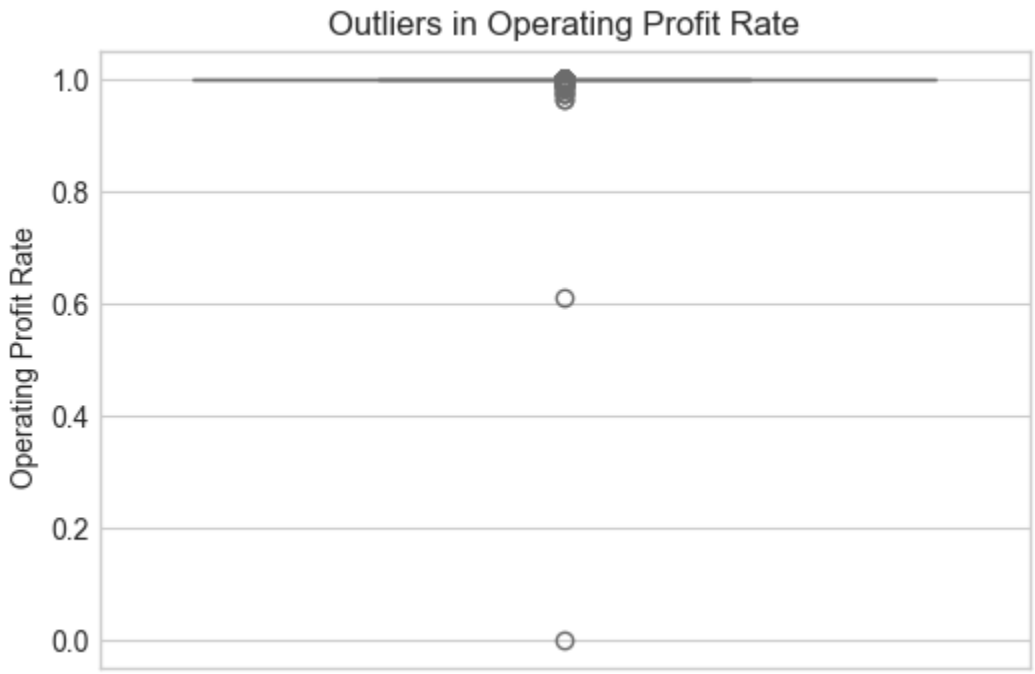


Feature Comparison by Bankruptcy Status

Outlier Analysis

Exploratory analysis further indicates the presence of extreme outliers in several financial ratios, particularly in leverage and profitability measures. While these outliers may reflect genuine financial distress, they can negatively affect certain machine learning algorithms.

Therefore, appropriate preprocessing steps—such as robust scaling or tree-based models that are less sensitive to outliers—are considered in subsequent modeling stages.



Outlier Visualization Using Box Plots for Selected Financial Ratios