ΕΛΛΗΝΙΚΗ
ΔΗΜΟΚΡΑΤΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΜΑΚΕΔΟΝΙΑΣ

University of Macedonia,Department of Applied Information Science and Computer Technology

Comparative Analysis of Classification Techniques for Bankruptcy Prediction: A Machine Learning Approach

Students

Kontaksis Ioannis

Ioannis Venetidis

Tsavalias Vasilios Ephraim

Abstract:

A critical component of risk management for creditors, investors, and legislators is identifying businesses that face bankruptcy. Timely interventions to mitigate potential losses and prevent systemic failures can be made possible by early detection of financial distress. With the ability to analyze financial data and make more accurate predictions about bankruptcy risk, machine learning has become a potent tool. To determine how well machine learning techniques can predict corporate bankruptcy, this study uses a wide range of techniques, such as Random Forests, K-Nearest Neighbors (KNN), Logistic Regression (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machines (SVM), and Cross-Validation. Based on the Bayes theorem, which divides data points into discrete categories according to their probability of belonging to each class, LDA and QDA are statistical classification techniques. QDA does not impose the assumption that the data points in each class follow a multivariate Gaussian distribution with a common covariance matrix, as does LDA. Data points are classified by KNN, a non-parametric classification algorithm, into the class most common among their closest neighbours. Although it is computationally efficient, the number of neighbours can have a significant impact.

A strong classification algorithm called Support Vector Machines (SVM) finds a hyperplane that efficiently divides data points into different classes. Although it can be computationally demanding, it is especially effective when handling non-linear data. Multiple decision trees are combined in Random Forests, an ensemble learning technique, to increase classification accuracy. It is an adaptable algorithm that can manage intricate feature interactions. To assess how well machine learning models perform, a technique called cross-validation separates the data into training and testing sets. The training set is used to train the model, and the testing set is used to evaluate the model's performance and generalizability. Preprocessing the data, choosing features, training the model, and assessing performance are all part of the study's methodology. ROC curve analysis, precision score, and F1 score are used to assess how well the different machine learning techniques perform. The findings show that compared to the other techniques, Random Forests and QDA perform better at predicting corporate bankruptcy. These results demonstrate how machine learning can improve the accuracy of bankruptcy predictions and support financial industry stakeholders in making well-informed decisions.

Table of Contents:

1. Theoretical Background:

In the realm of bankruptcy prediction, a critical task in financial risk management, the evolution of modelling approaches has been significantly shaped by the challenges inherent in financial data. This journey from traditional statistical methods to advanced machine learning techniques reflects the evolving complexities of financial data analysis.

Evolution of Modeling Approaches:
Traditional Statistical Methods: Initially, techniques like logistic regression and discriminant analysis were prevalent. These methods, while foundational, often faltered in capturing complex, non-linear relationships present in financial datasets. To address these limitations, machine learning models such as Support Vector Machines (SVMs), Random Forests, Neural Networks, and Gradient Boosting Engines have risen to prominence. These models excel in handling large datasets and unravelling intricate patterns.

Challenges and Solutions in Bankruptcy Prediction:

- Class Imbalance: Bankruptcy cases are rarer than non-bankruptcy ones, leading to potential model bias. Techniques like SMOTE and custom loss functions are employed to address this imbalance.

- Feature Selection: Critical in enhancing model performance. Methods like backward feature elimination and Lasso regression aid in identifying the most predictive variables, reducing complexity, and improving interpretability.

Key Metrics to Optimize:

Accuracy: Indicates the overall correctness of the model. In bankruptcy prediction, high accuracy ensures the model reliably identifies both healthy and bankrupt firms.

Precision: Measures the accuracy of bankruptcy predictions. High precision is vital to avoid mislabeling healthy firms as bankrupt, thereby preventing unnecessary interventions.

Recall (Sensitivity): Assesses the model's ability to detect all relevant cases, particularly firms that are likely to go bankrupt. High recall is crucial for capturing most firms at risk, and facilitating timely interventions.

F1 Score: The F1 score is the harmonic mean of precision and recall, effectively balancing these two metrics A high F1 score is critical in bankruptcy prediction for several reasons. A robust F1 score indicates that the model adeptly balances these aspects, avoiding both unnecessary interventions in healthy firms and missing out on identifying firms genuinely at risk.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve): This metric measures a classifier’s ability to differentiate between classes (bankrupt versus non-bankrupt firms) across various thresholds. It's integral to bankruptcy prediction, as it evaluates the model's performance across a range of thresholds, providing a comprehensive

view of its predictive power. This is crucial in bankruptcy prediction, where the cost associated with false positives and false negatives can be substantial.

2. Introduction:

In this thesis, the focus is on the application of machine learning in predicting corporate bankruptcy, an area gaining significant attention in financial analysis and risk management. The core idea is that the effectiveness of bankruptcy prediction depends not only on the algorithms used but also on the quality and relevance of the financial data.

The study examines various machine learning models, including Logistic Regression, Decision Trees, Random Forest, and Quadratic Discriminant Analysis (QDA), assessing their predictive abilities in bankruptcy detection.

An important methodological aspect is the use of fold-stratified cross-validation, essential for addressing the imbalance in bankruptcy datasets, where instances of failure often outnumber those of financial stability. This approach ensures each data subset accurately reflects the entire dataset, enhancing the reliability of model evaluations.

The thesis also investigates performance metrics like precision, accuracy, recall, F1 score, and area under the receiver operating characteristic curve (AUC ROC). These metrics offer a comprehensive assessment of model performance, surpassing the limitations of single-metric evaluations.

The objective is to identify effective machine learning techniques for bankruptcy prediction and to enhance understanding of their optimal application. Through in-depth analysis and comparison of models, this research contributes to financial risk assessment and supports more informed decision-making in the business world.

The paper employs various machine learning techniques to predict company bankruptcy. Methods like Logistic Regression, Quadratic Discriminant Analysis (QDA), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forests, and Decision Trees are evaluated for their bankruptcy prediction capabilities.

LDA and QDA, classification algorithms, use Bayes' theorem to classify data into classes. LDA assumes data from each class is drawn from a Gaussian distribution with a shared covariance matrix, while QDA does not assume a common covariance matrix.

KNN classifies data points based on the majority vote of nearest neighbors, without assuming any specific data distribution. Its simplicity is offset by sensitivity to the choice of neighbours.

SVM classifies data points into classes using a hyperplane and is effective even with non-linearly separable data, showcasing strong classification abilities. The inclusion of Support Vector Machines (SVM) in the thesis recognizes its robust classification capabilities but also notes its computational demands, especially with large datasets.

Random Forests, an ensemble learning method, combines several decision trees to enhance classification accuracy. This algorithm is versatile and effective for diverse classification problems, but the performance can be influenced by the number of trees in the forest.

Cross-validation, a key technique for assessing machine learning algorithms, involves splitting data into training and testing sets. The algorithm is trained on the training set and evaluated on the test set. This method is crucial for determining an algorithm's generalization performance.

By evaluating the effectiveness of these methods in predicting bankruptcy, the research aims to provide valuable insights for policymakers, creditors, and investors, aiding them in making informed decisions about business strategies and investments.

The F1 score, which combines precision and recall, is a metric for evaluating model accuracy. Precision measures the proportion of correct positive predictions, while recall indicates the proportion of actual positives correctly identified. The F1 score, which balances the importance of precision and recall, ranges from 0 (no accurate predictions) to 1 (perfect accuracy).

Precision score, another metric, reflects the accuracy of positive predictions. A higher precision score implies fewer false positives.

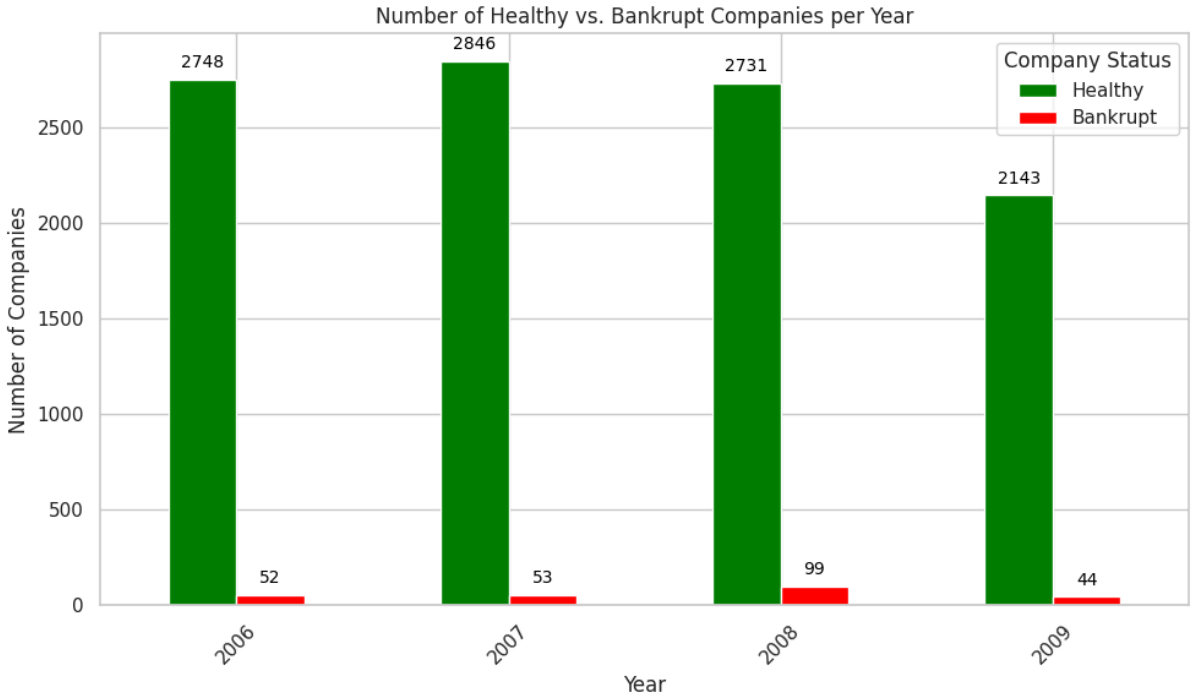
The ROC curve, a graphical representation, shows a model's performance at different thresholds. It plots the true positive rate (TPR) against the false positive rate (FPR). TPR measures the proportion of actual positives correctly identified, while FPR indicates the rate of false positives. The ROC curve is valuable for evaluating a model's performance across various discrimination levels.

3. Methodology:

The following section will analyze the data set used
including sources and key characteristics (performance indicators, binary activity indicators, company status and year). Then will follow the pre-processing steps (data pre-processing), such as the handling of missing values and data normalization. And finally, the following classification models will be implemented, evaluated and optimized:
Knn, Decision Trees, Random Forest, Naive Bayes, Logistic Regression, SVM, LDA, and QDA (which were chosen by the team members) in turn they will be evaluated through the measurements: Accuracy, Precision, Recall, F1 score, Area under the ROC curve (AUC-ROC).

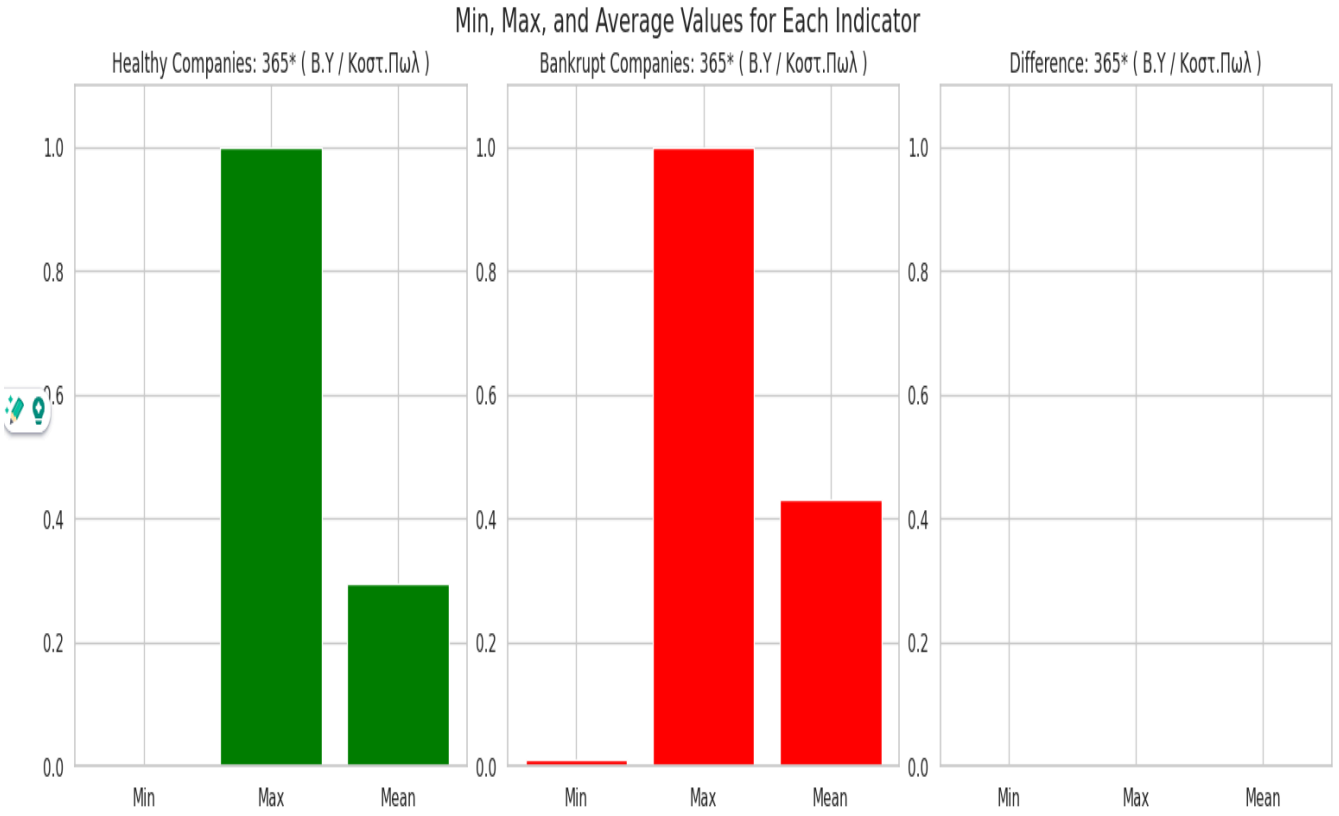
3.1 Overview and pre-processing of Data:

Loading the dataset into a pandas DataFrame is a crucial step, as it facilitates data manipulation and analysis in Python. The use of data.head() for initial exploration helps in understanding the dataset's structure by displaying a snapshot of its contents.
The dataset undergoes a process to check for missing values, ensuring data integrity and completeness. This step is essential for accurate analysis. Additionally, the data is normalized using a min-max scaling technique. This normalization excludes certain columns like 'INCONSISTENCY INDICATION (=2) (v+1)' and 'YEAR', likely due to their unique data properties or relevance in the analysis.
The objective is to employ a data visualization tool that effectively communicates the historical distribution of successful and unsuccessful companies. Presenting the data in an annotated bar graph format is chosen for its ability to quickly convey trends and patterns over time. This visualization method is not only insightful for data analysis but also instrumental in decision-making processes, particularly in evaluating a company's financial health. By offering a clear visual representation, it allows users to easily discern patterns and make informed assessments.



The observation that a substantial majority of companies in the dataset are performing well highlights the presence of an unbalanced dataset. This imbalance emphasizes the need for the f1 metric, which is particularly adept at handling datasets where one class is significantly underrepresented.

Another crucial element in this analysis is the comparative examination of key statistics between financially stable and unstable companies. This comparison yields insights into how these groups differ across various indicators. Visualizations play a vital role in this context. They aid in identifying patterns and anomalies in the data, which can be pivotal in guiding decision-making processes and further in-depth analysis. Presenting a sample visualization is a practical approach, especially when the full chart is too extensive to include in its entirety in a handout. This strategy allows for a focused examination of critical data points, providing a representative overview of the broader dataset.



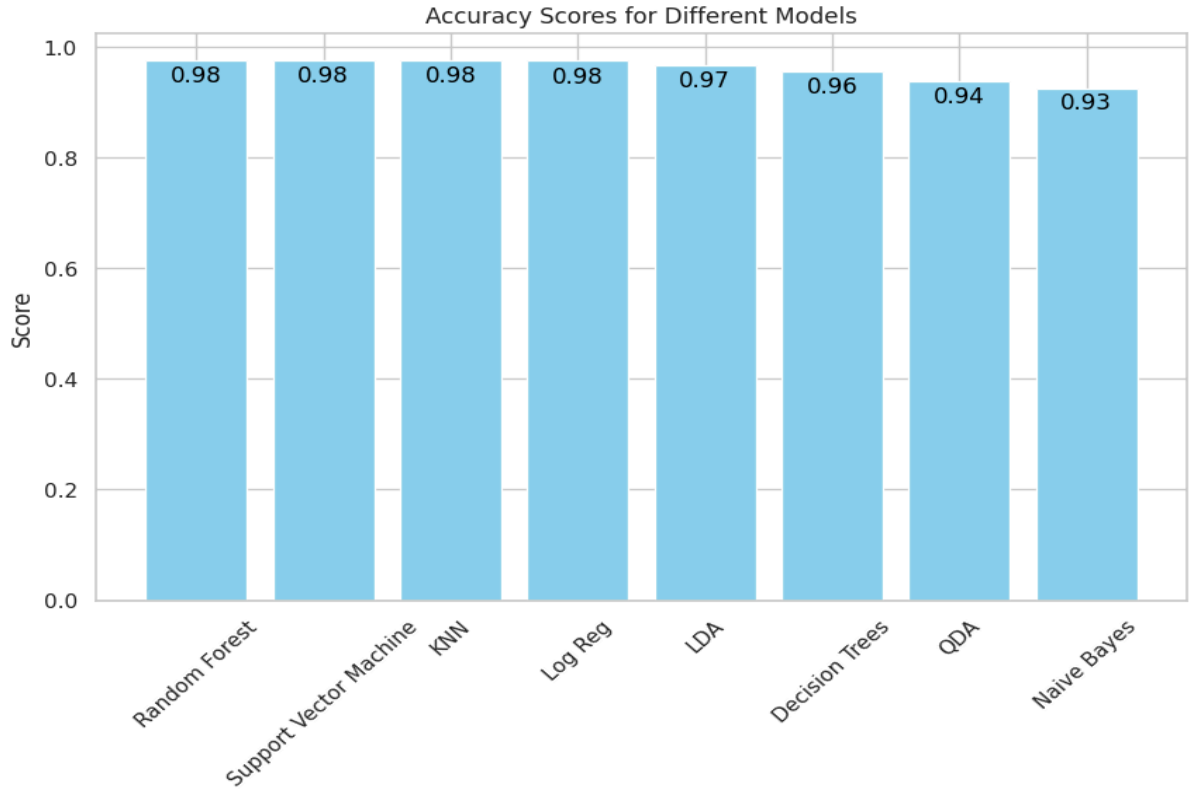
- Graph One - Average Revenue/Cost of Sales: This graph indicates that on average, bankrupt companies exhibit a lower revenue-to-cost of sales ratio compared to healthy companies. This trend suggests that bankrupt companies are generally less efficient in generating revenue relative to their costs. It reflects a potential operational inefficiency or a challenging market position that hampers their ability to cover costs through revenue.

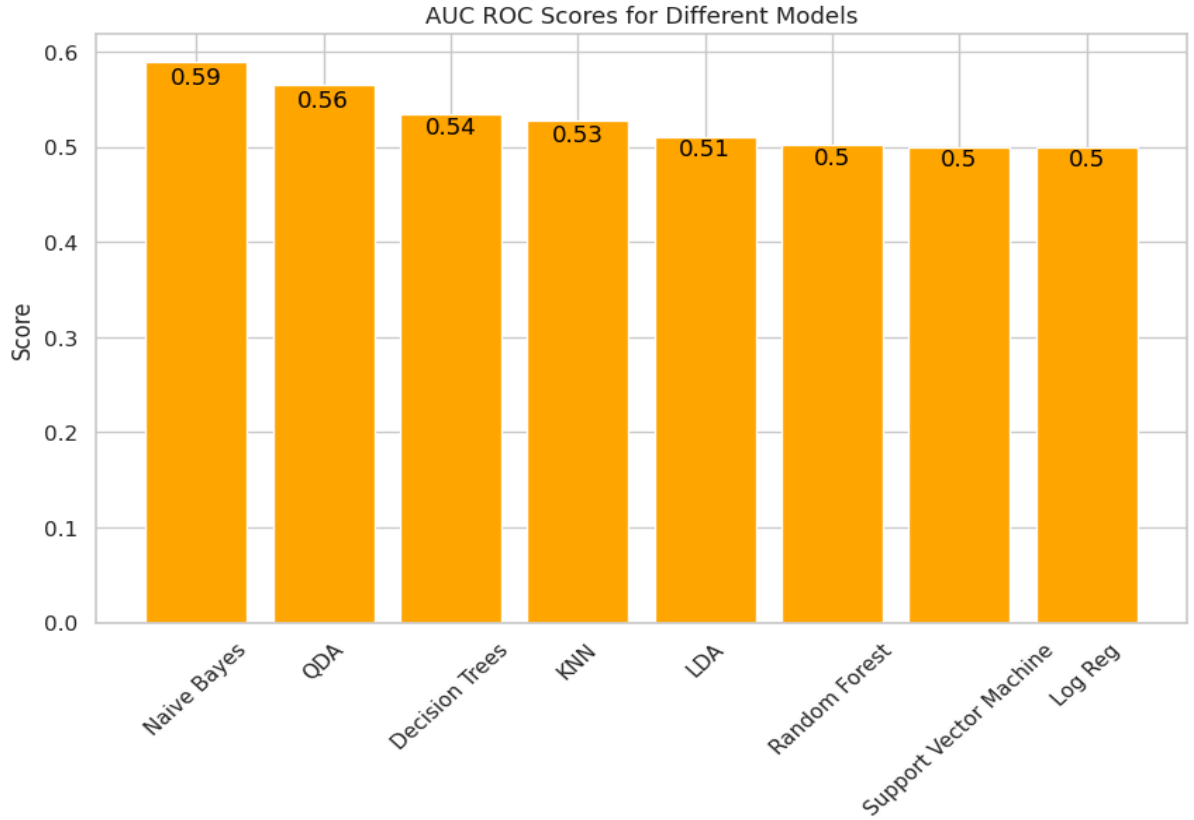
- Graph Two - Maximum Revenue/Cost of Sales: Contrasting with the first graph, this one shows that the maximum ratio for bankrupt companies surpasses that of healthy ones. This implies that while the average trend points to lower efficiency among bankrupt companies, there are exceptions where some bankrupt companies achieved high revenue levels. However, these high revenues did not necessarily translate to financial stability, possibly due to proportionally high costs.

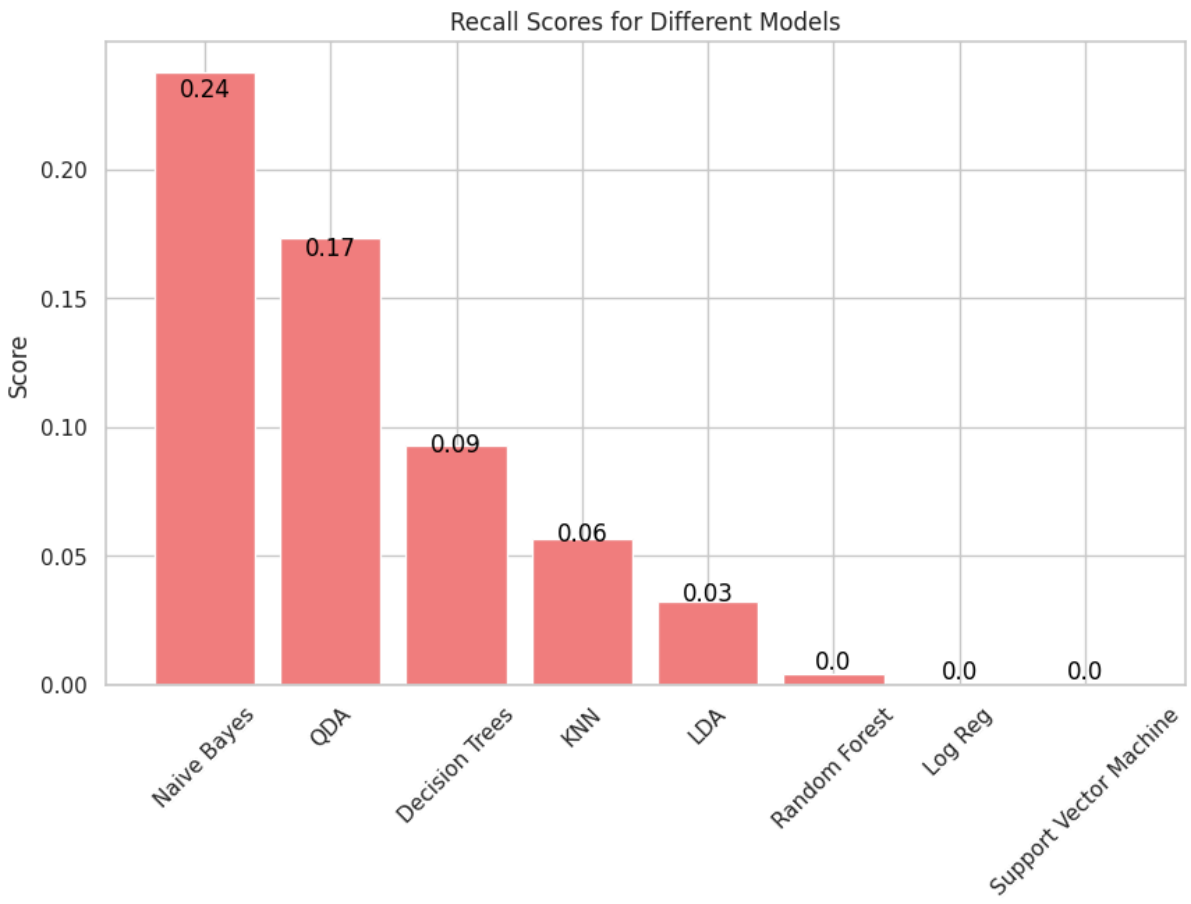
- Graph Three - Difference in Ratio: This chart illustrates the consistent positive difference in the revenue/cost of sales ratio between bankrupt and healthy firms across all index values. This positive difference underscores that bankrupt companies, on the whole, have a consistently lower ratio than their healthy counterparts.

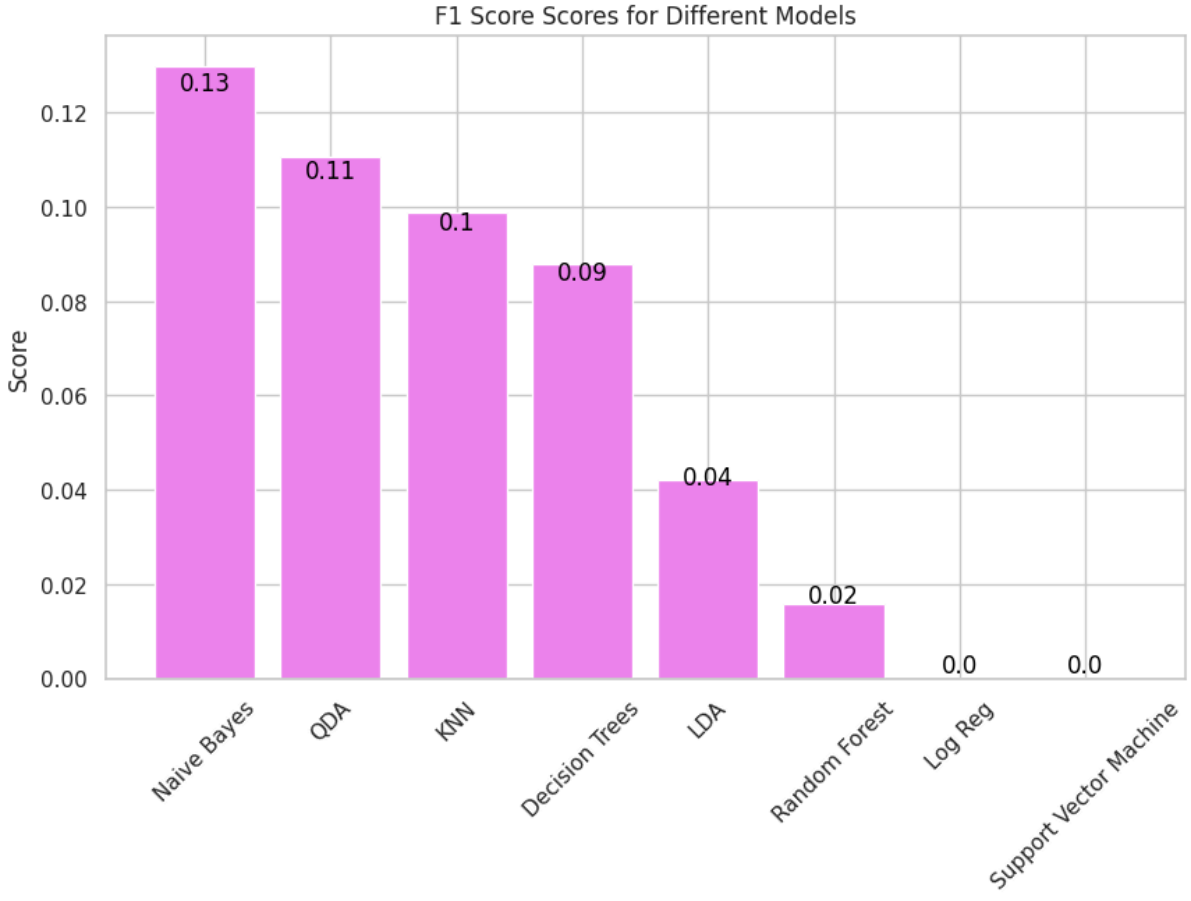
From these observations, it becomes evident that the revenue/cost of sales ratio is a significant metric in differentiating bankrupt from healthy companies. While bankrupt companies generally have a lower average ratio, indicating less efficiency in managing revenues against costs, there are instances where they demonstrate high revenue generation.

The subsequent part of the analysis will involve an initial assessment of the quality of various models and metrics, conducted without any fine-tuning or optimization. This preliminary evaluation will provide a baseline understanding of the models' performance in their default configurations.



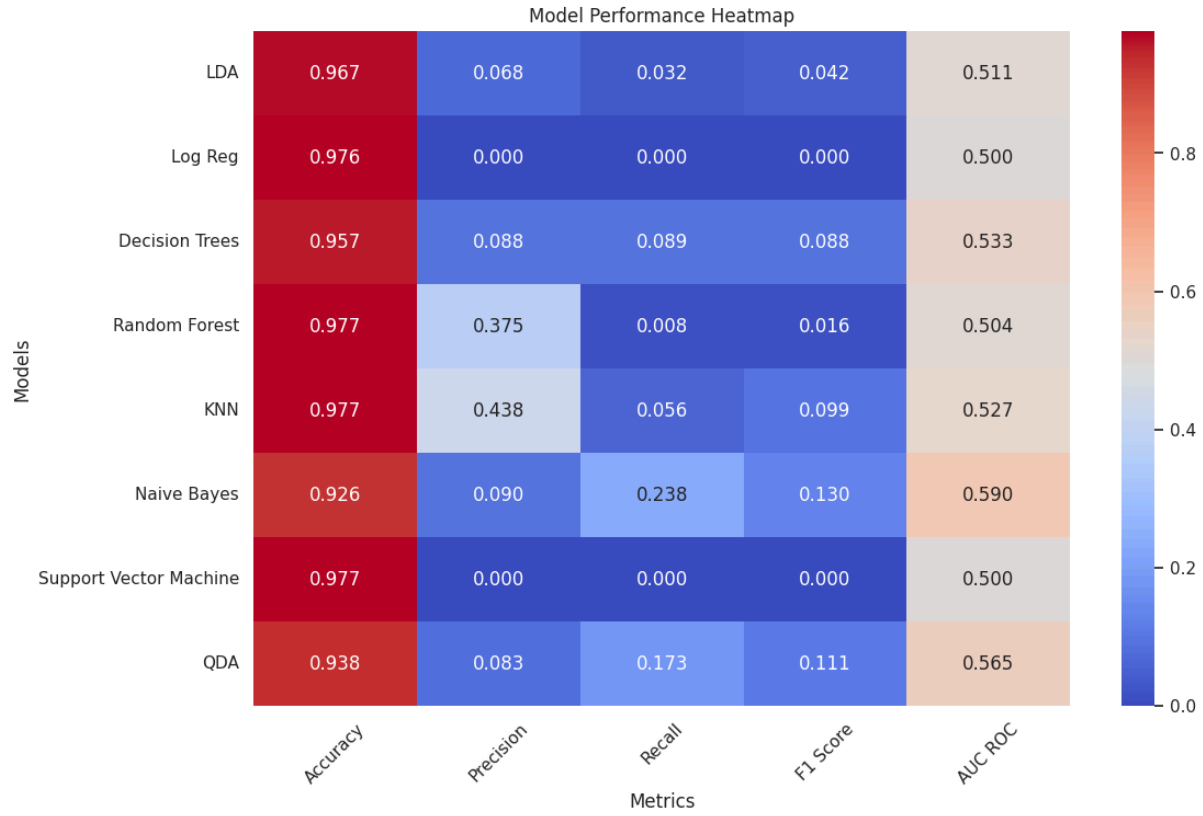



Recall Scores for Different Models
Score
0.00
0.05
0.10
0.15
0.20
0.24
0.17
0.09
0.06
0.03
0.0
0.0
0.0
Naive Bayes
QDA
Decision Trees
KNN
LDA
Random Forest
Log Reg
Support Vector Machine


F1 Score Scores for Different Models
Score
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.13
0.11
0.1
0.09
0.04
0.02
0.0
0.0
Naive Bayes
QDA
KNN
Decision Trees
LDA
Random Forest
Log Reg
Support Vector Machine

The initial evaluation of unprocessed machine learning techniques reveals their limited utility in their raw form, particularly when dealing with complex criteria. This is evident when comparing different performance metrics. For instance, while accuracy appears similar across all models, the recall metric displays significant variability.

Integrating the information from the earlier graphs into a heatmap provides a comprehensive model performance map. This visualization facilitates a clearer understanding of how various machine learning models perform across different metrics in the dataset.



The heatmap indicates that the Support Vector Machine (SVM) and Random Forest models show the most promising performance across all considered metrics. Conversely, the Naive Bayes model exhibits the weakest performance. However, it's premature to designate SVM as the superior model solely based on this data. The subsequent analysis emphasizes the pivotal role of data processing in selecting the optimal model for bankruptcy prediction.
In the context of bankruptcy prediction, achieving high accuracy is crucial to avoid unnecessary actions against financially stable firms.

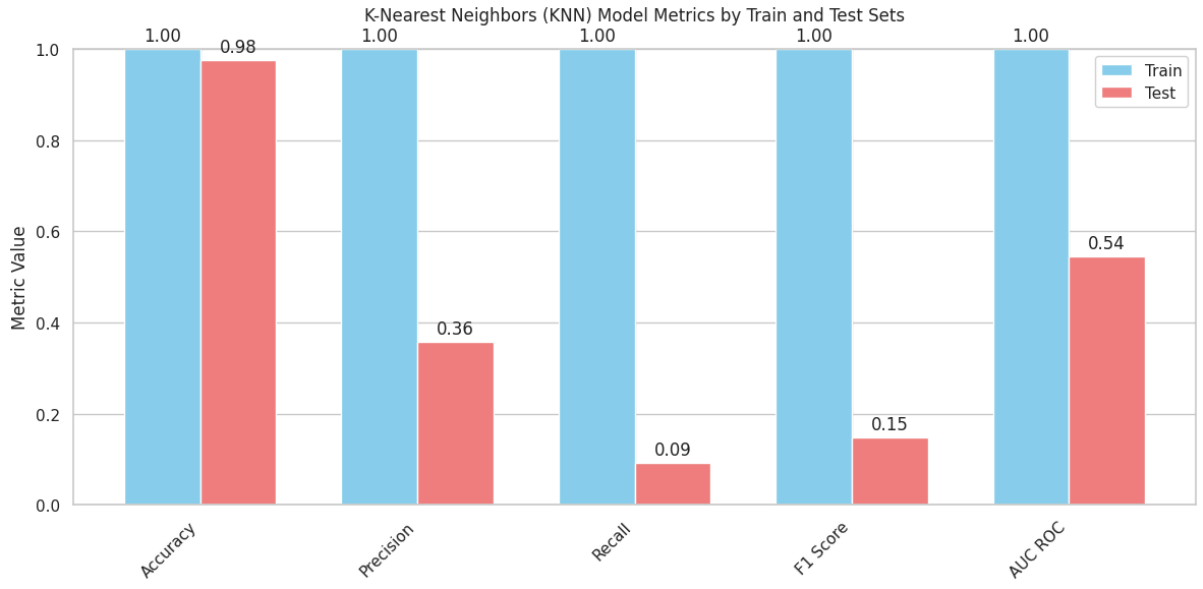
Simultaneously, a high recall is vital to ensure that most firms at risk of bankruptcy are correctly identified for potential intervention. Metrics like the F1 Score and AUC ROC offer a more nuanced perspective of a model's performance than mere accuracy. This depth of analysis is critical in bankruptcy prediction, where the consequences of misclassifying firms can be substantial. These metrics collectively provide a balanced view of a model's ability to accurately identify both bankrupt and healthy firms, an essential consideration in financial risk management.

3.2 Model Tuning:

The analysis underscores the necessity of focusing not only on the training set but also on the test set, striving to optimize them to enhance the performance metrics. This comprehensive approach is essential for developing robust machine-learning models capable of reliable predictions.
When evaluating each model, it is crucial to consider the balance between Type I errors (false positives) and Type II errors (false negatives). This balance directly impacts business decisions:

- False Positives: Incorrectly identifying healthy businesses as at risk can lead to unwarranted interventions, potentially disrupting stable operations.
- False Negatives: Failing to identify businesses that are genuinely on the brink of bankruptcy could mean missing critical opportunities for timely support or intervention.

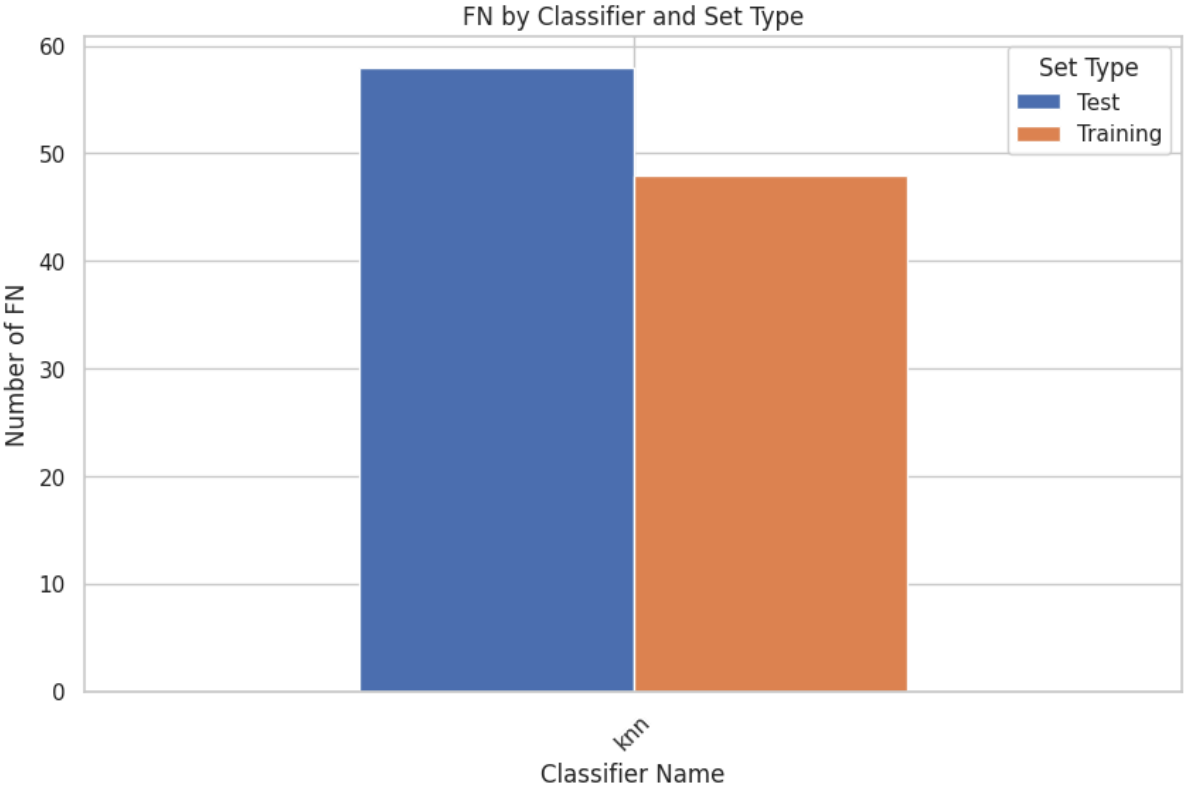
3.2.1 K-Nearest Neighbors (Knn):

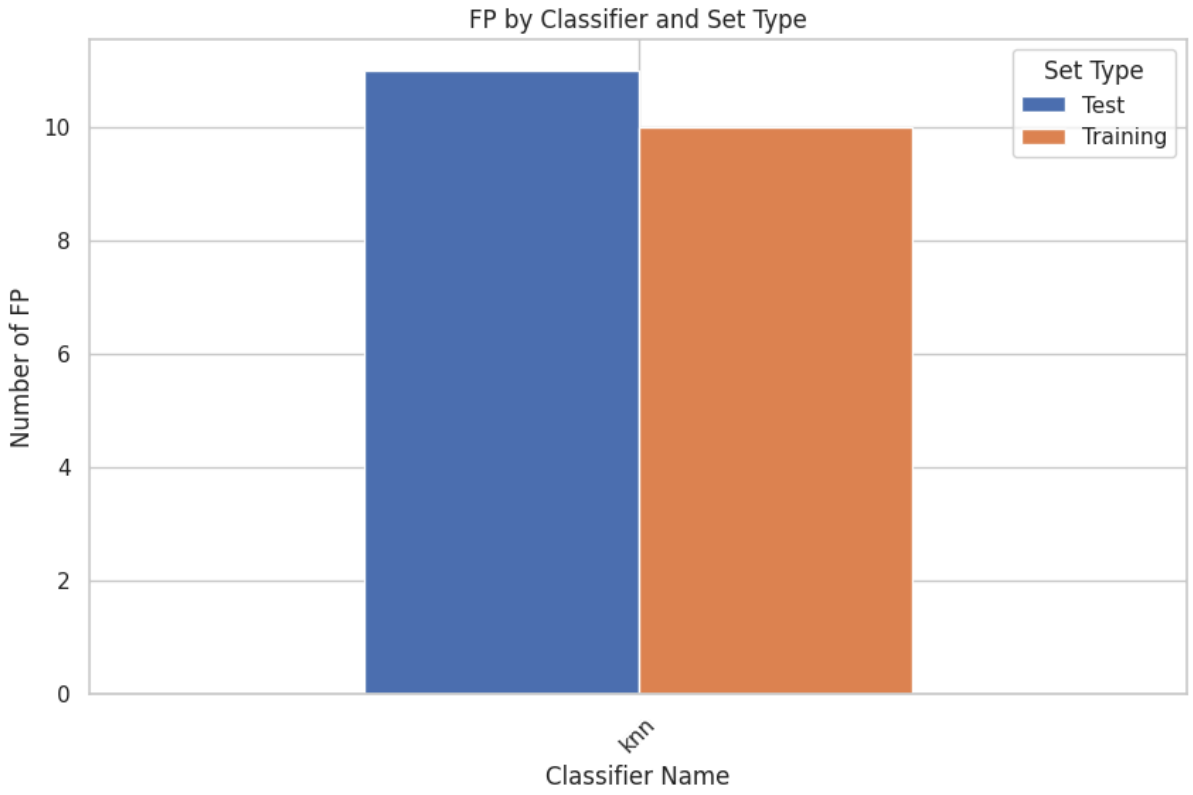


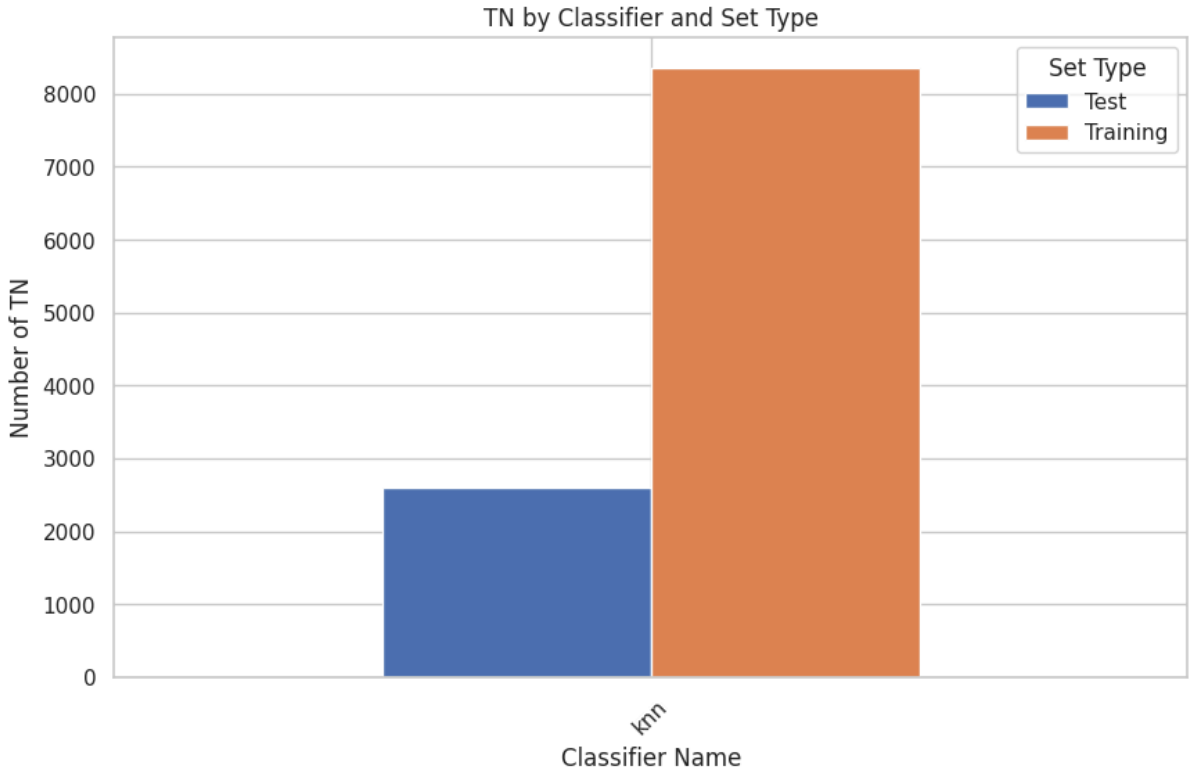
- Accuracy: While high accuracy scores in both training and testing are promising, perfect scores in the training set could be indicative of overfitting. This means the model may be too closely tailored to the training data, limiting its applicability to new, unseen data.
- Precision, Recall, F1-Score: The significant disparity between the training and testing metrics suggests that the model may not generalize well. Generalization is key for models to be effective in real-world applications, where data can vary from the training set.

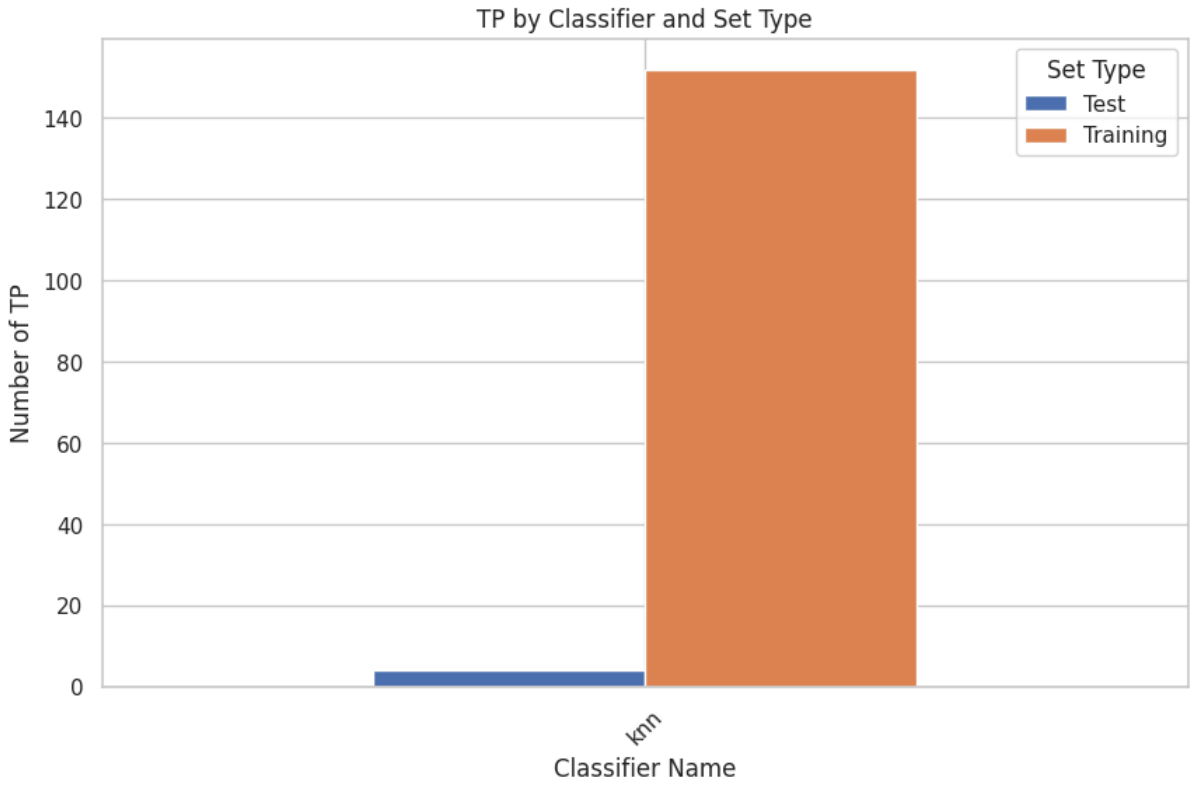
- AUC-ROC: Achieving perfect scores in the training set but only average in the testing set could suggest that the model is overfitting. AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is a critical metric for evaluating a model's ability to distinguish between classes, and a significant drop in performance from training to testing warrants attention.

Regarding the KNN (K-Nearest Neighbors) model, while it shows perfect performance on the training set, this excellence does not extend to the testing set, where the metrics are noticeably lower. This discrepancy could result from noise within the dataset or a lack of representative examples in the training set. In the context of business health predictions, it's vital that the model not only memorizes the training data but also generalizes effectively to new and varied data. This generalization is essential for the model to be reliable in making predictions about business outcomes in different and potentially unforeseen scenarios.

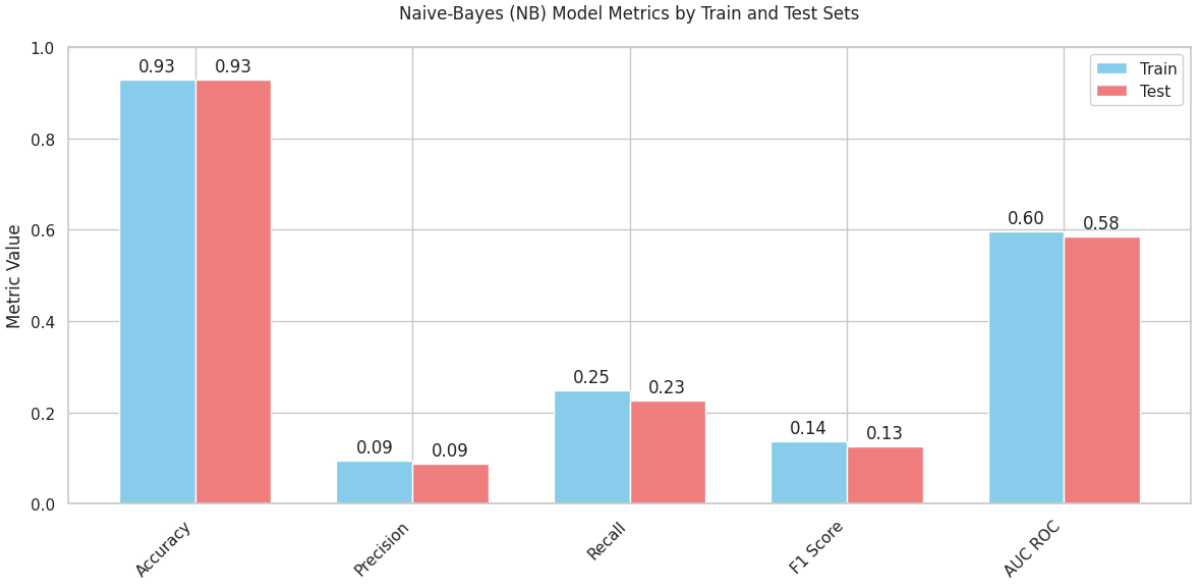



FP by Classifier and Set Type
Set Type
Test
Training
Number of FP
0
2
4
6
8
10
knn
Classifier Name


TN by Classifier and Set Type
Set Type
Test
Training
Number of TN
0
1000
2000
3000
4000
5000
6000
7000
8000
knn
Classifier Name



3.2.2 Naive Bayes(Nb):

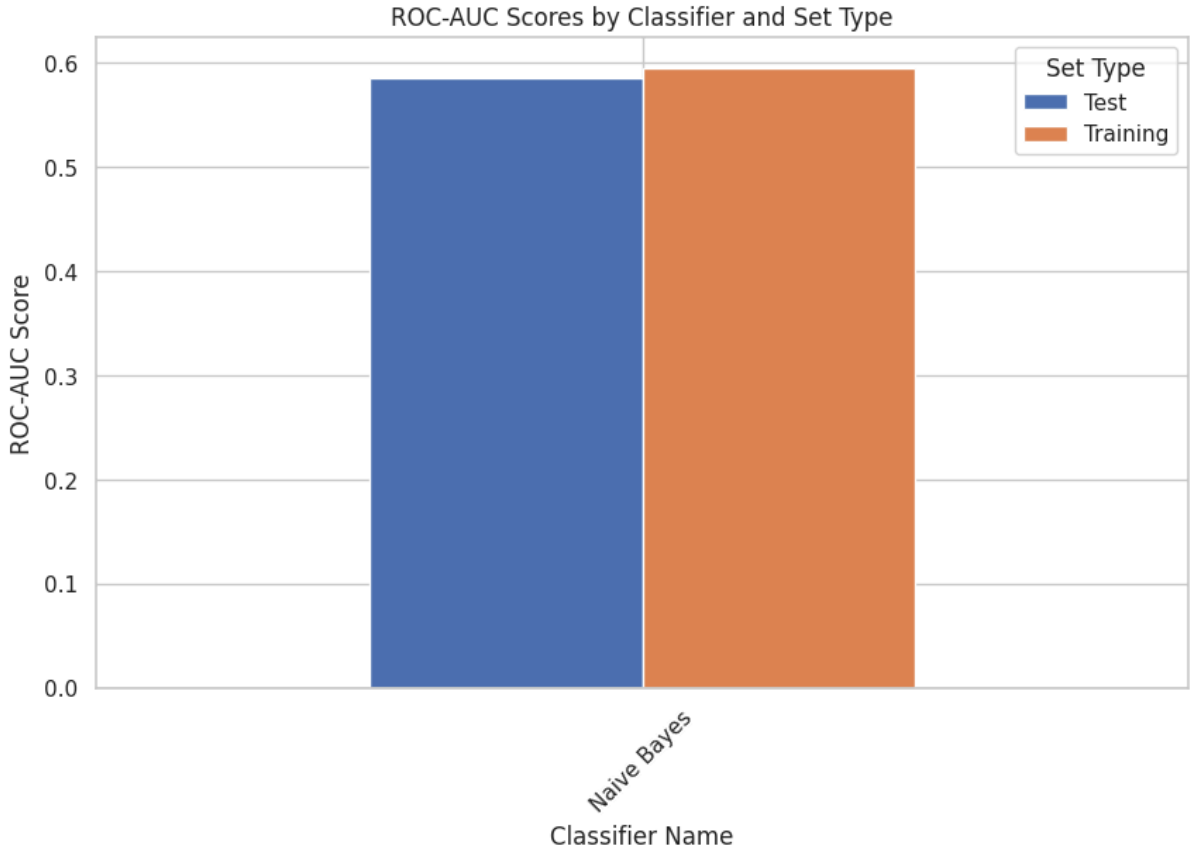


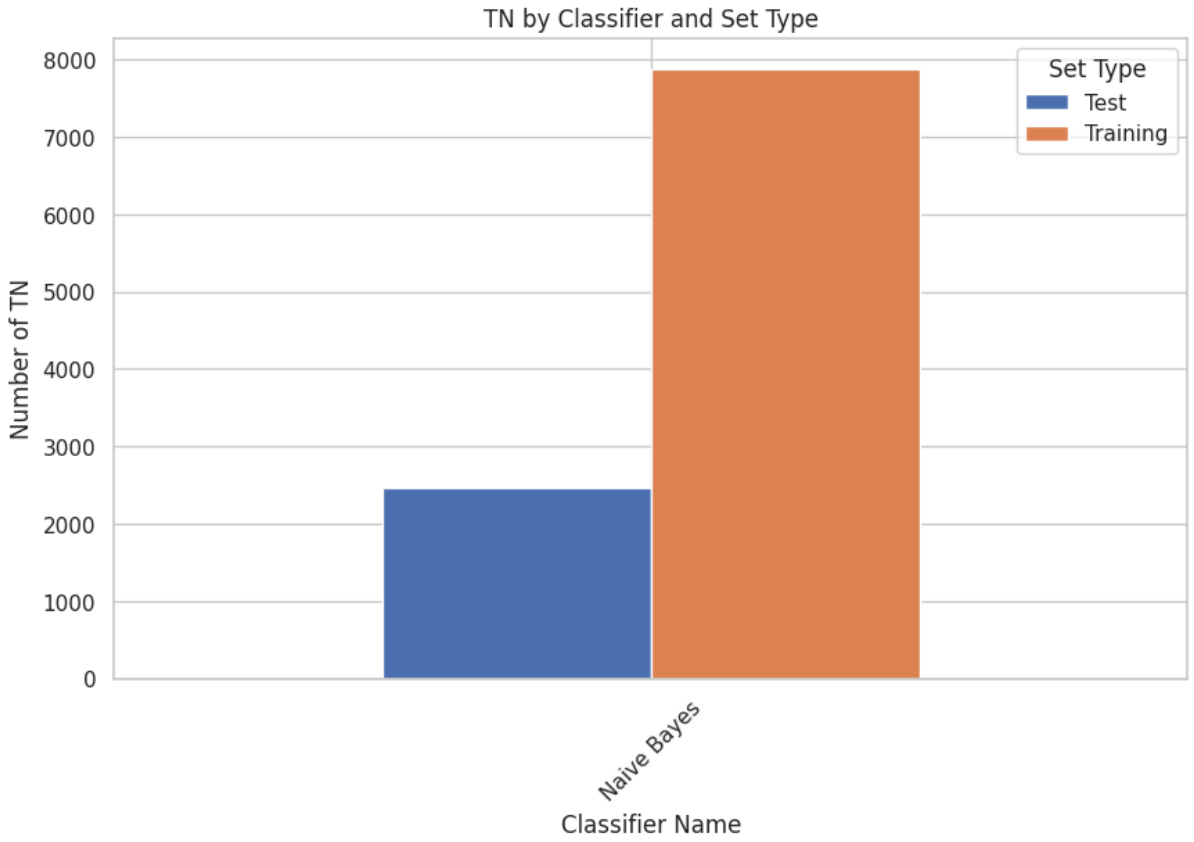
Naive Bayes Model:

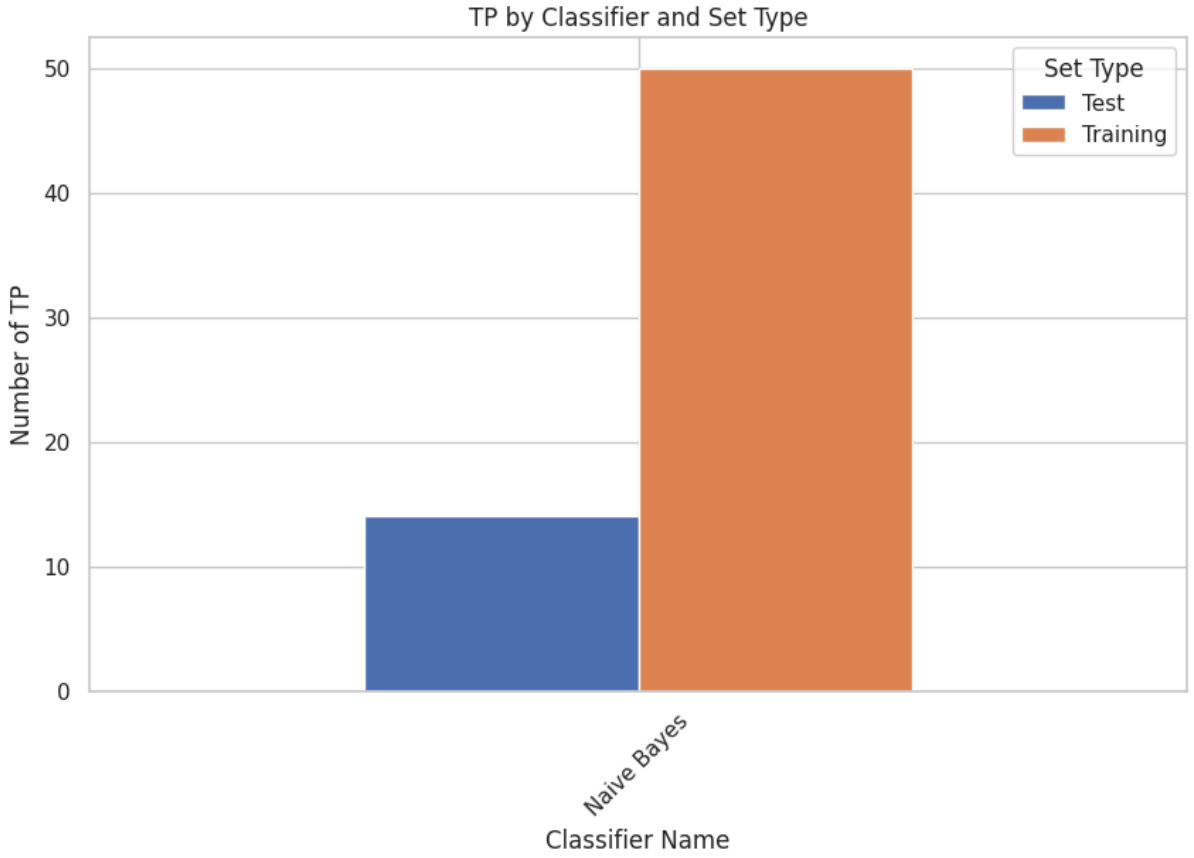
- Accuracy: Although Naive Bayes shows decent accuracy, the concern lies in its precision, particularly in the testing set. The model's tendency towards high false positive rates is significant.

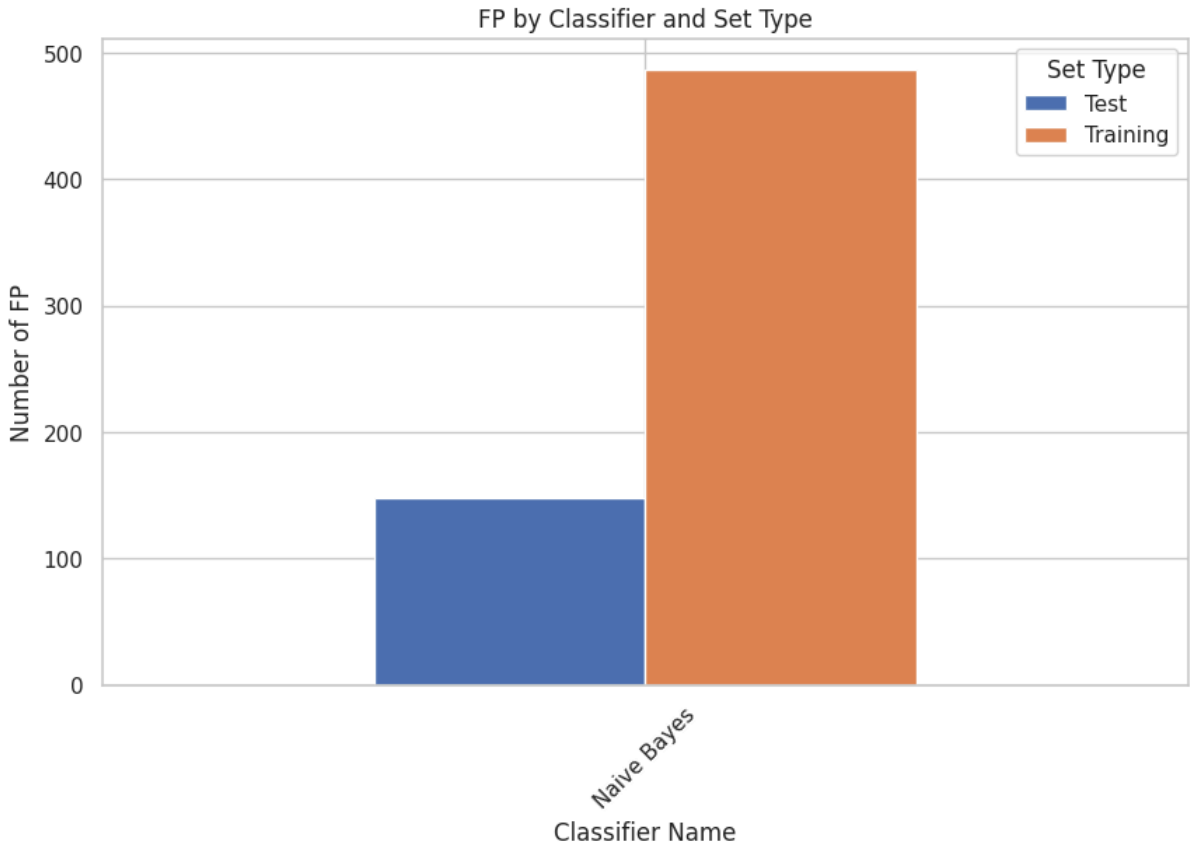
- Precision, Recall, F1-Score: All these metrics are low, with precision being notably poor in both training and testing sets. This suggests a high rate of false positives, where the model predicts bankruptcy where it does not exist.
- AUC-ROC: The average AUC-ROC score indicates a moderate ability to discriminate between classes. While the model can distinguish between healthy and failing firms to some extent, its overall performance is unremarkable.

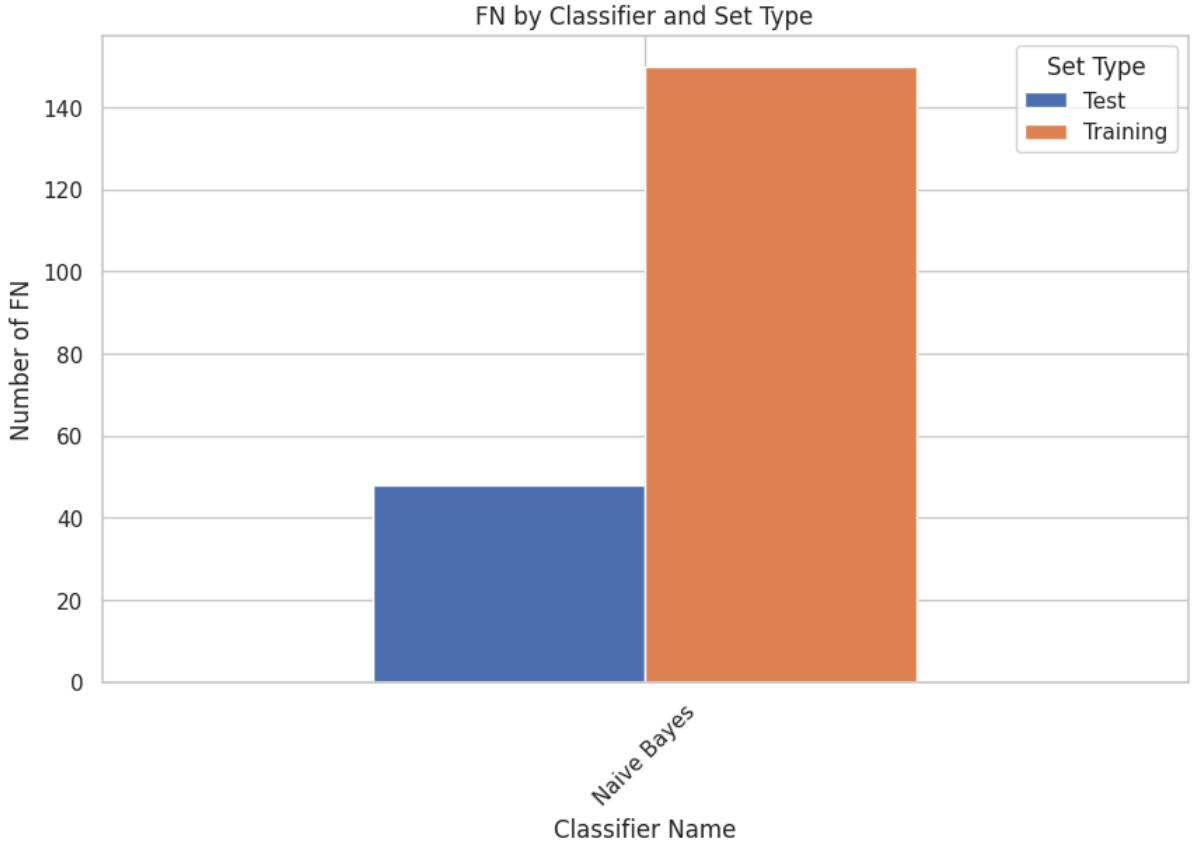
The simplicity of Naive Bayes, often an advantage in certain contexts, may be a drawback here, failing to capture the complexities inherent in business health indicators. The risk of false alarms is a serious concern, potentially leading to unnecessary interventions.



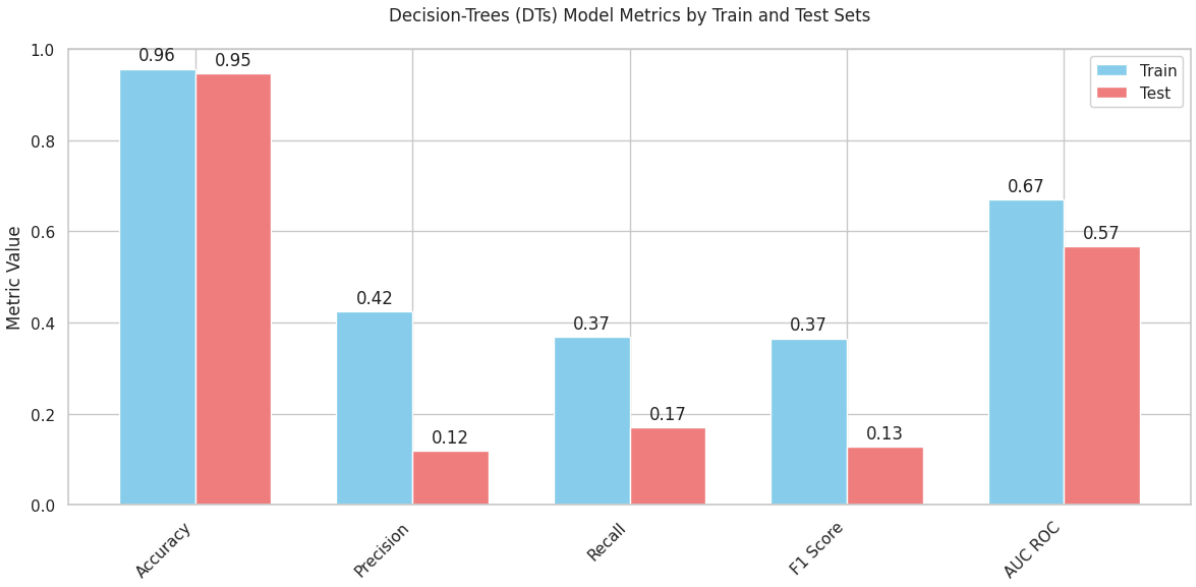

TN by Classifier and Set Type
Set Type
Test
Training
Number of TN
8000
7000
6000
5000
4000
3000
2000
1000
0
Naive Bayes
Classifier Name


TP by Classifier and Set Type
Set Type
Test
Training
Number of TP
50
40
30
20
10
0
Naive Bayes
Classifier Name





3.2.3 Decision Trees(Dts):

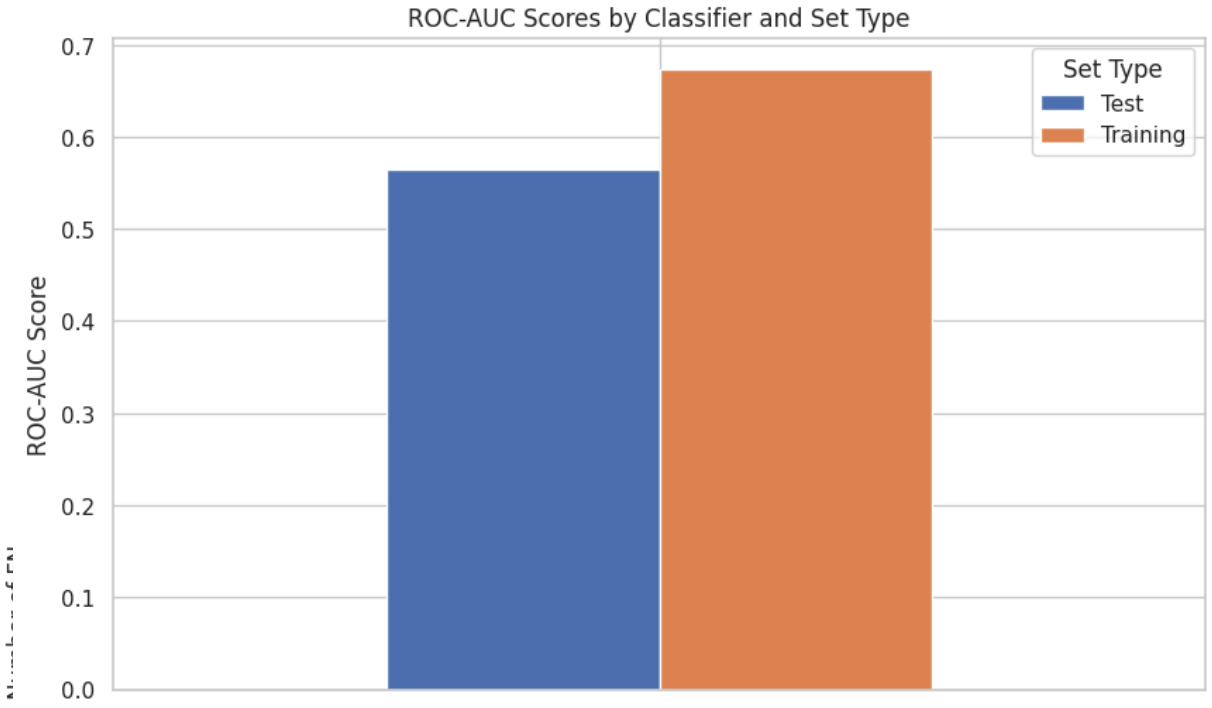


- Accuracy: There is a notable disparity between training and testing accuracy, suggesting overfitting. The model may be too closely tailored to the training data, compromising its effectiveness on new data.

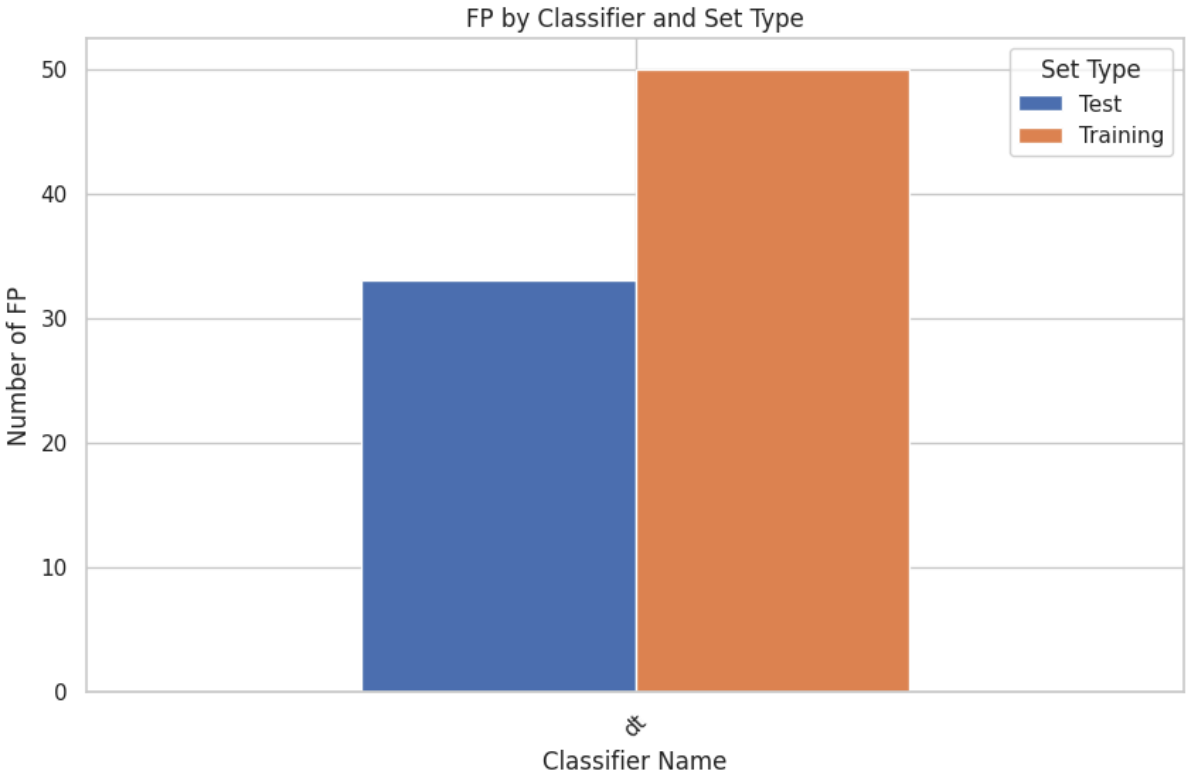
- Precision, Recall, F1-Score: These metrics are low in the testing set, indicating the model is prone to both false positives and false negatives. This is concerning, as it means the model often incorrectly predicts the financial health of firms.

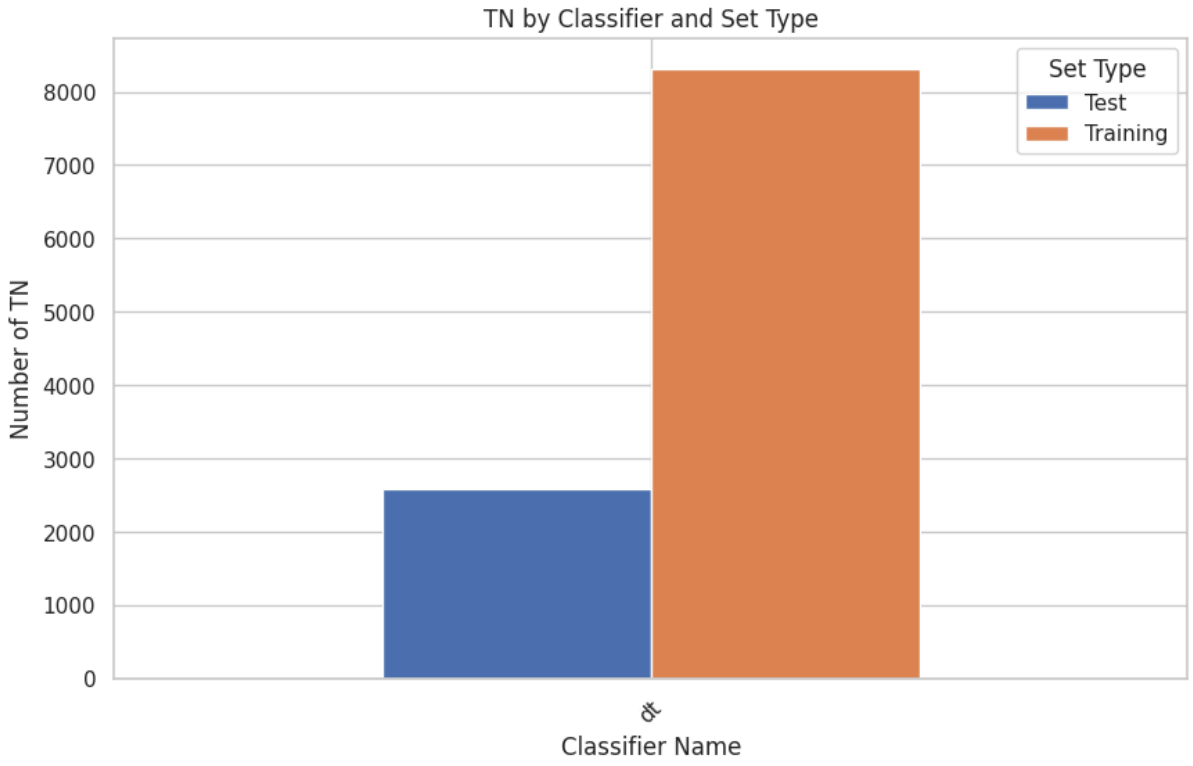
- AUC-ROC: Moderate AUC-ROC values further suggest the model's limited ability to differentiate between healthy and bankrupt companies.

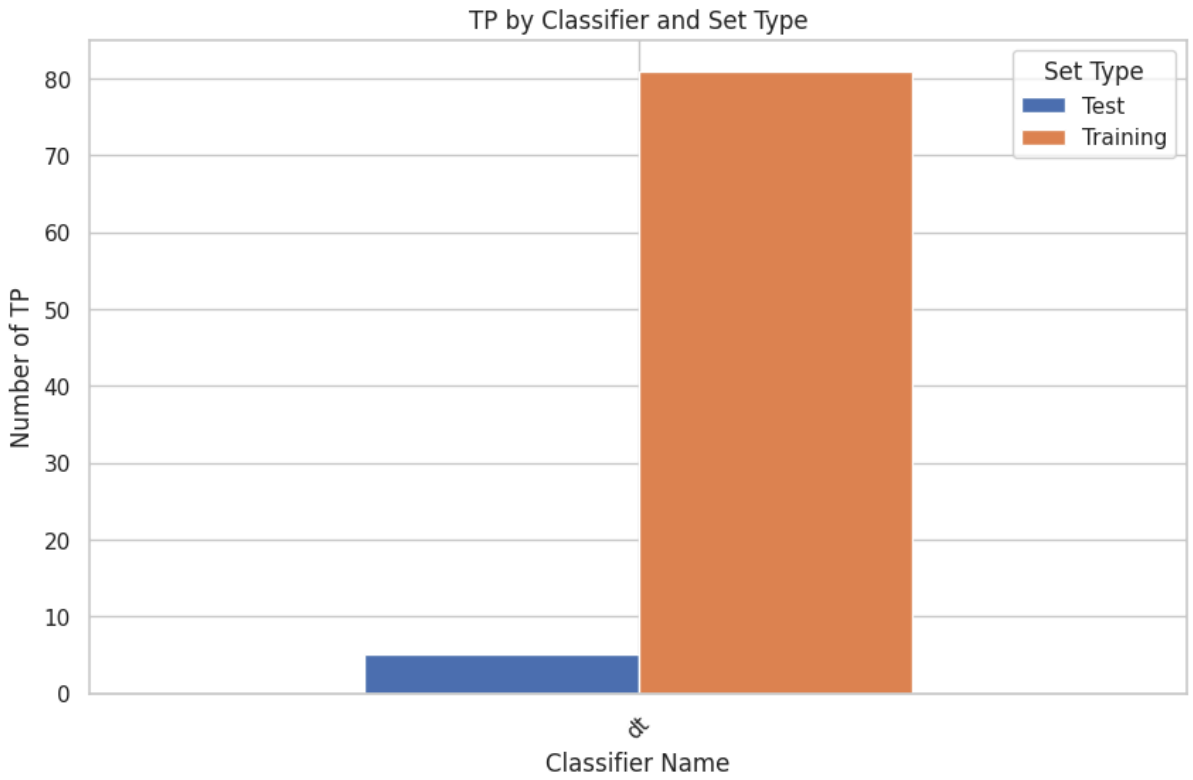
The substantial performance drop from training to testing implies the model memorizes training data rather than learning generalizable patterns. This is problematic for practical applications where accurate classification of a firm’s financial status is critical. Adjusting the model's complexity or employing pruning techniques might enhance its generalizability and reduce overfitting.


ROC-AUC Scores by Classifier and Set Type
Set Type
Test
Training
ROC-AUC Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
dt
Classifier Name

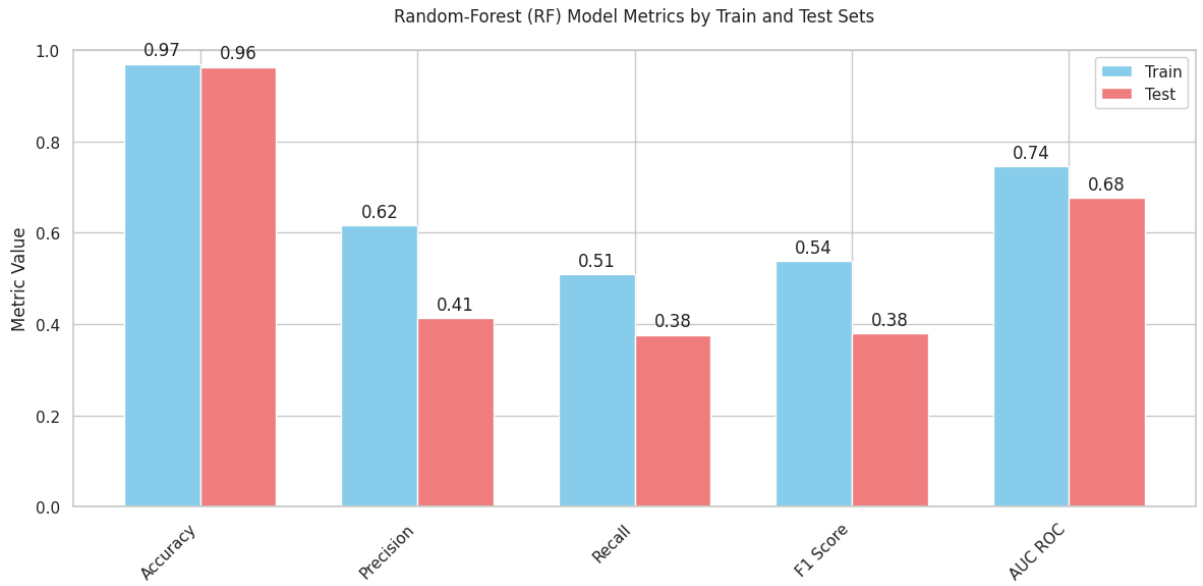

0
20
dt
Classifier Name


FP by Classifier and Set Type
Set Type
Test
Training
Number of FP
0
10
20
30
40
50
dt
Classifier Name


TN by Classifier and Set Type
Set Type
Test
Training
Number of TN
8000
7000
6000
5000
4000
3000
2000
1000
0
dt
Classifier Name


TP by Classifier and Set Type
Set Type
Test
Training
Number of TP
80
70
60
50
40
30
20
10
0
dt
Classifier Name

3.2.4 Random Forest(RF):

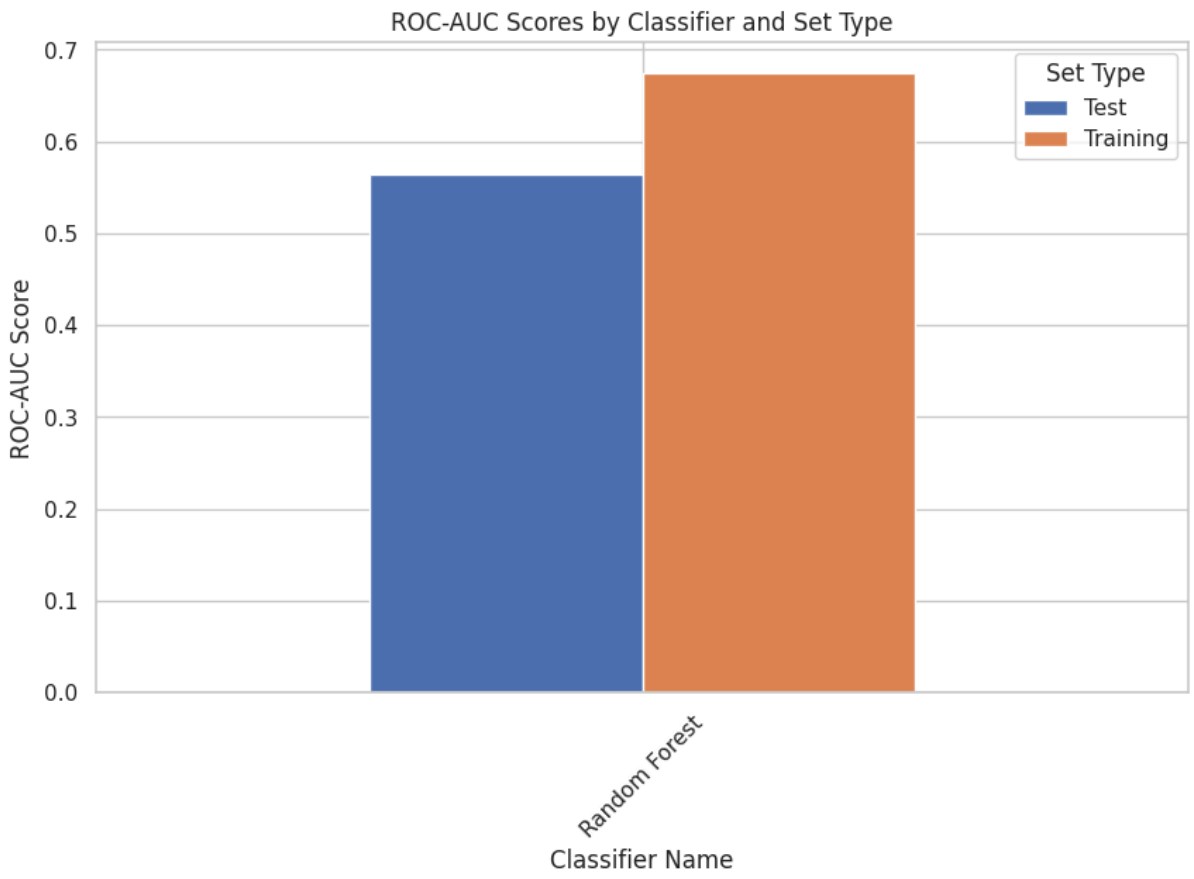


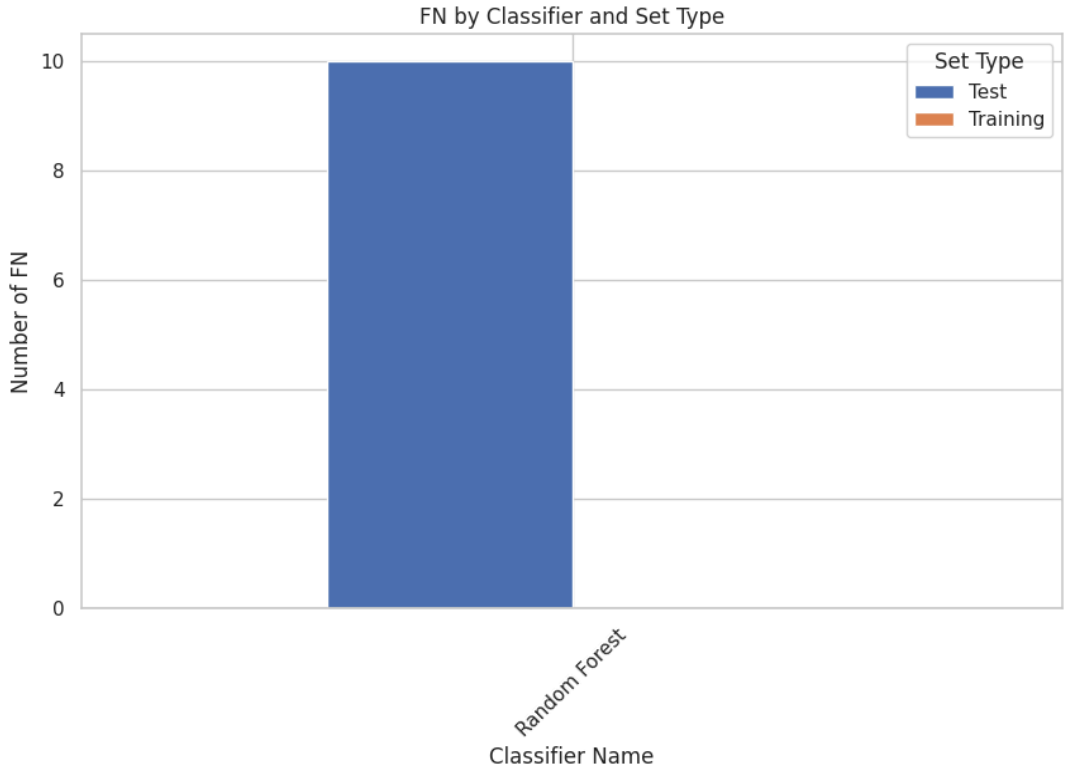
- Accuracy: Exhibits very high accuracy in both training and testing sets, which is a strong indicator of the model's robustness.

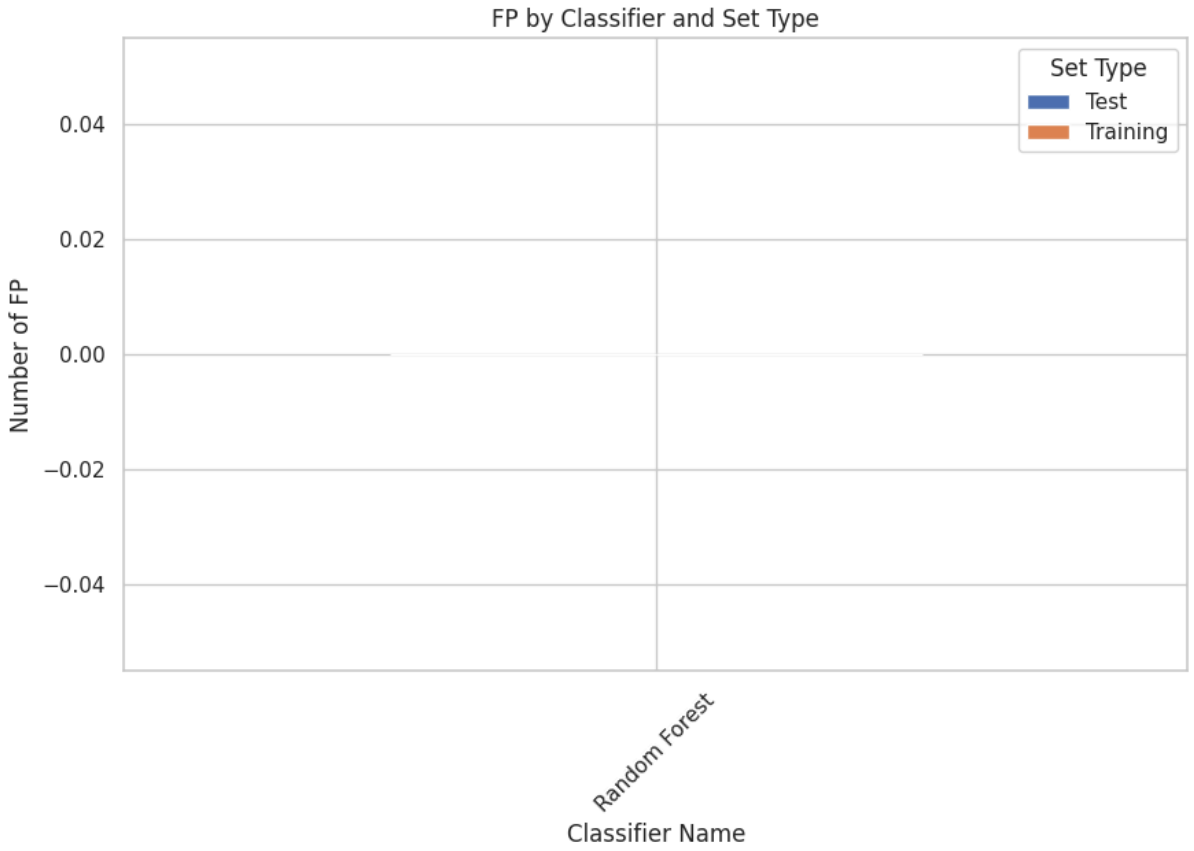
- Precision, Recall, F1-Score: These metrics are better compared to Decision Trees, which is expected since a Random Forest is essentially an aggregation of Decision Trees. This ensemble approach helps in addressing the overfitting issues often seen in individual Decision Trees.

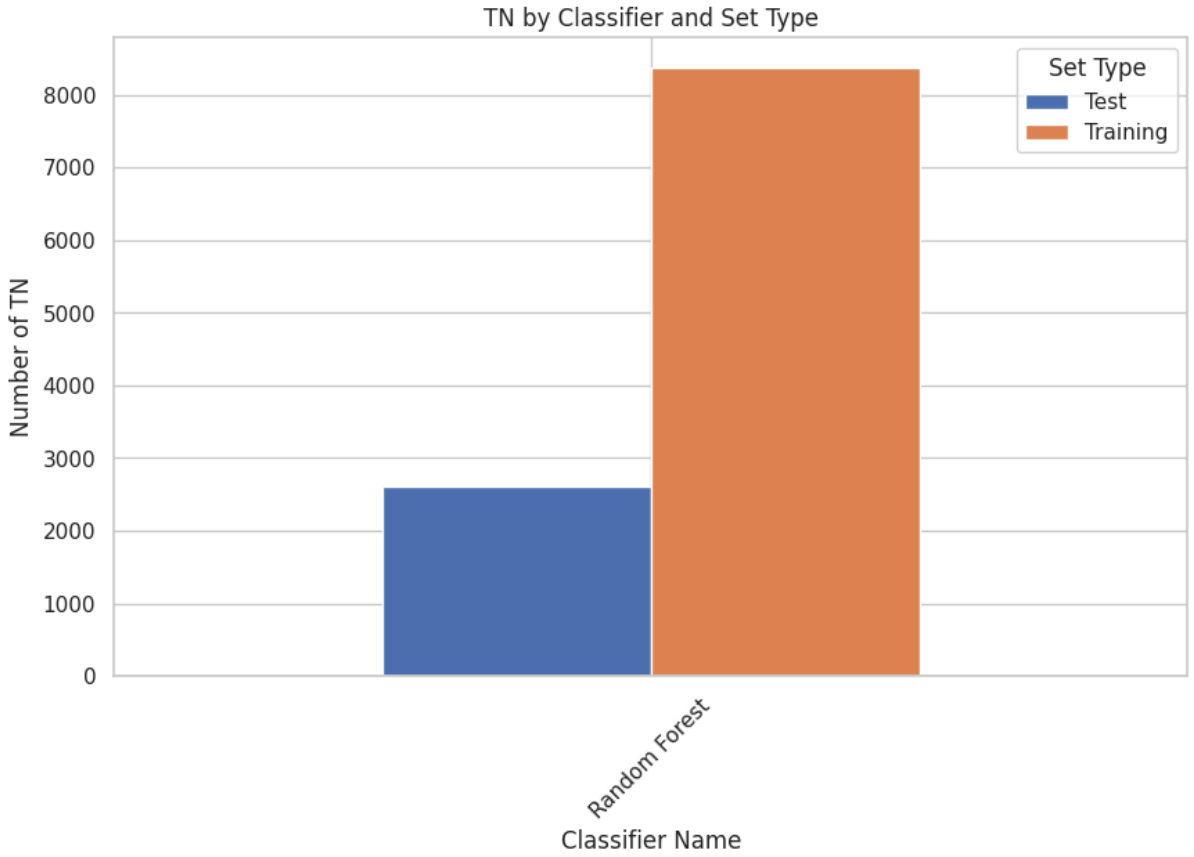
- AUC-ROC: The high AUC-ROC values are indicative of the model's strong classification capabilities. This metric reflects the model's ability to differentiate between healthy and bankrupt firms effectively.

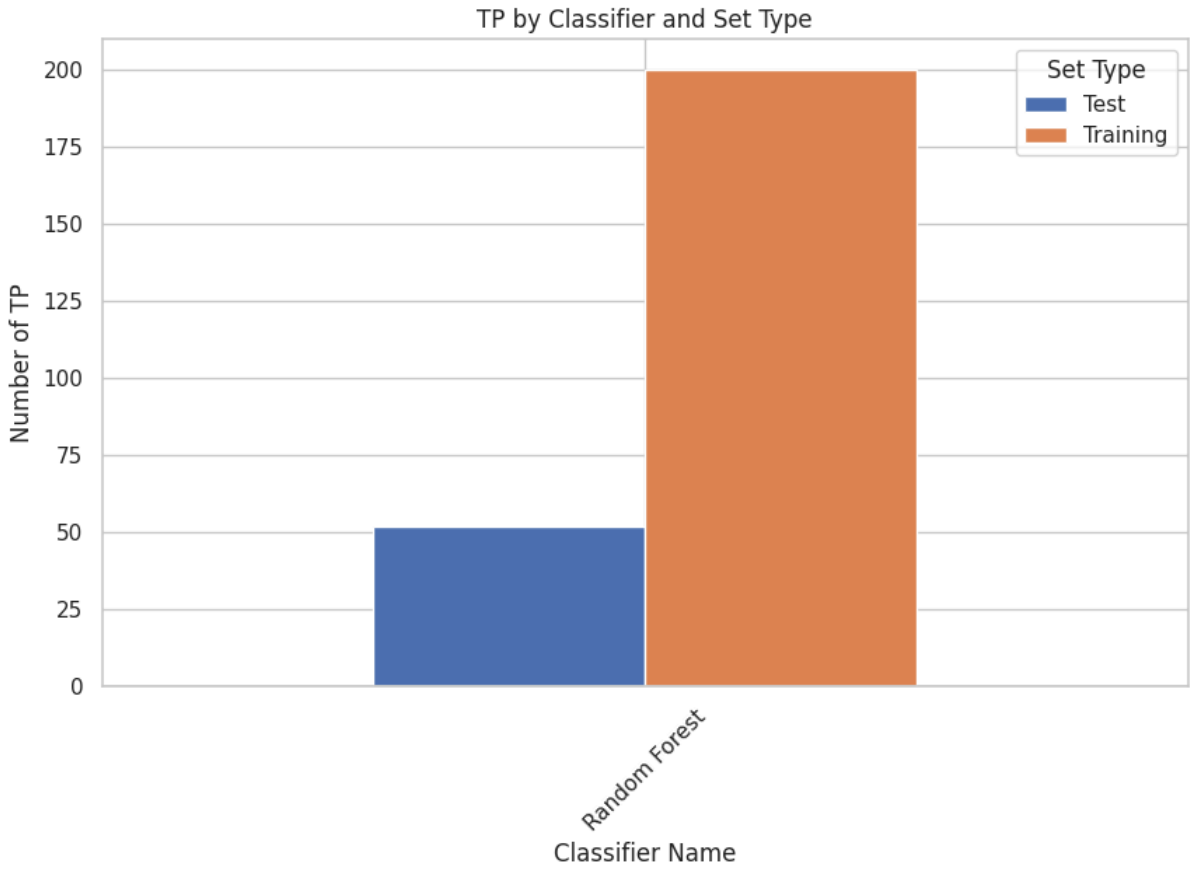
The overall strong performance across all metrics suggests that the Random Forest model effectively captures underlying patterns in the data, potentially making it a valuable tool for identifying firms at risk of bankruptcy. Its ability to handle overfitting enhances its reliability for practical applications.


ROC-AUC Scores by Classifier and Set Type
Set Type
Test
Training
ROC-AUC Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Random Forest
Classifier Name


FN by Classifier and Set Type
Set Type
Test
Training
Number of FN
0
2
4
6
8
10
Random Forest
Classifier Name


FP by Classifier and Set Type
Set Type
Test
Training
0.04
0.02
0.00
−0.02
−0.04
Number of FP
Random Forest
Classifier Name


TN by Classifier and Set Type
Set Type
Test
Training
8000
7000
6000
5000
4000
3000
2000
1000
0
Number of TN
Random Forest
Classifier Name



3.2.5 Linear Discriminant Analysis(LDA):

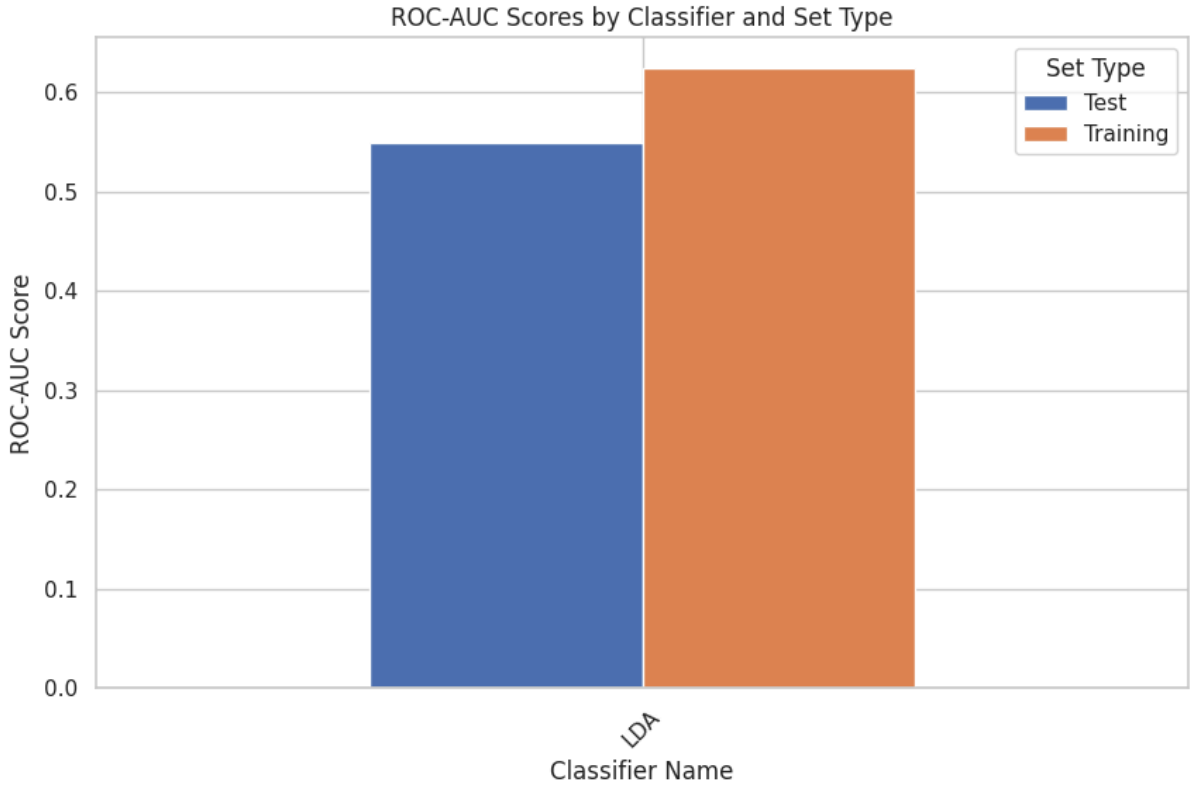
- Accuracy: Shows high accuracy in both training and testing sets, indicating its effectiveness in capturing linear boundaries between healthy and failed firms.

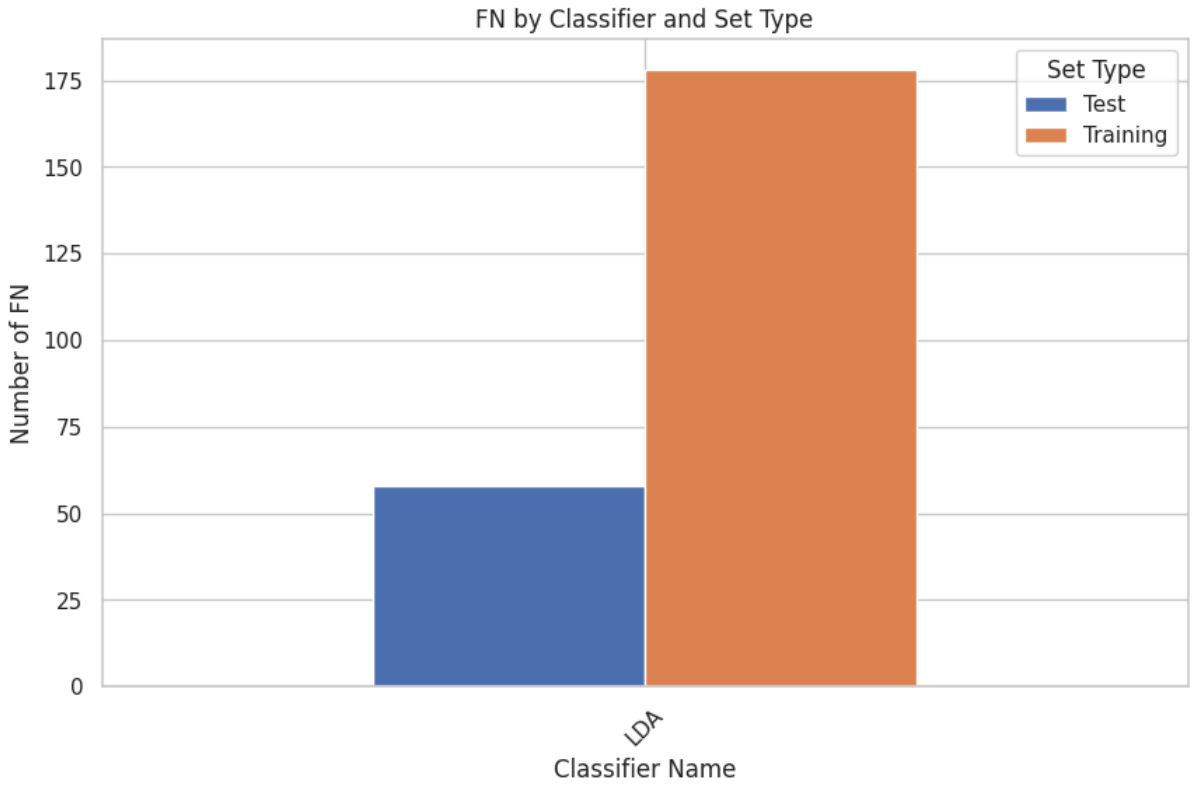
- Precision and Recall: Both metrics are moderate with a noticeable drop in the testing set, suggesting the presence of some false positives and false negatives.

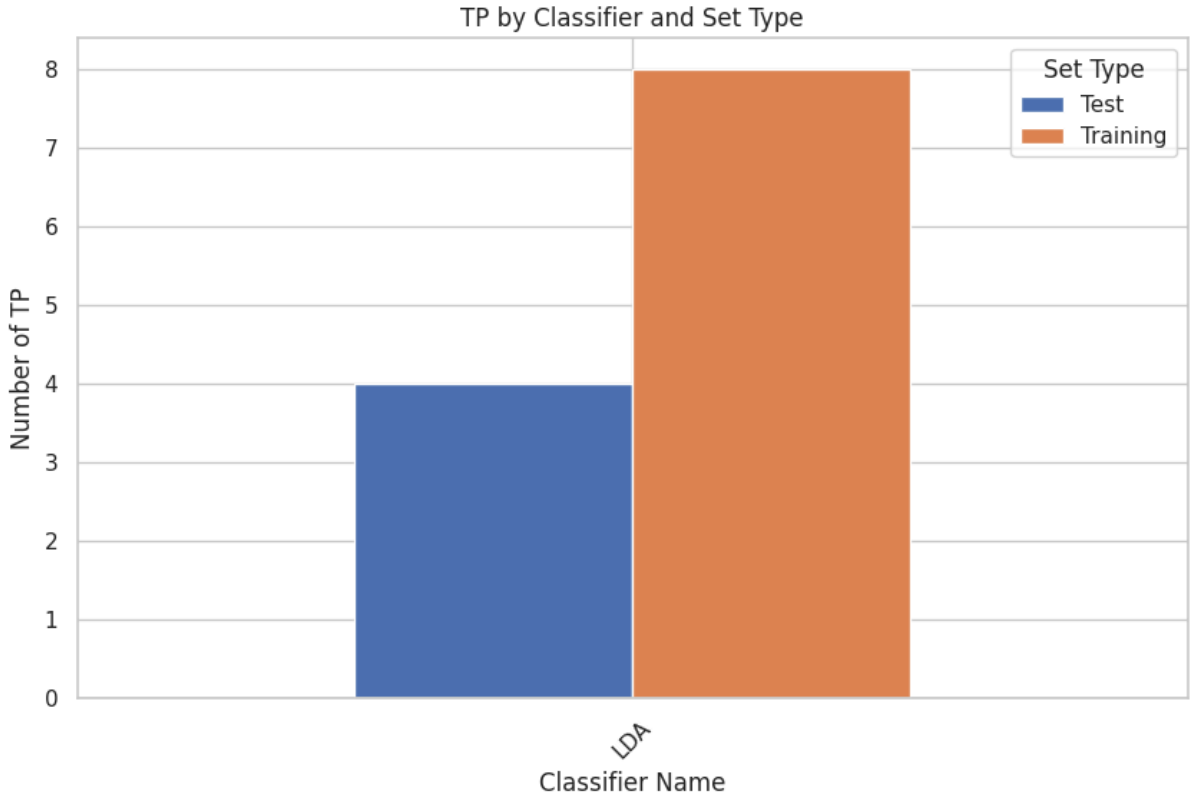
- F1 Score: The moderate F1 score implies a balance between precision and recall, though it is not optimal. This balance is important in assessing the overall effectiveness of the model.

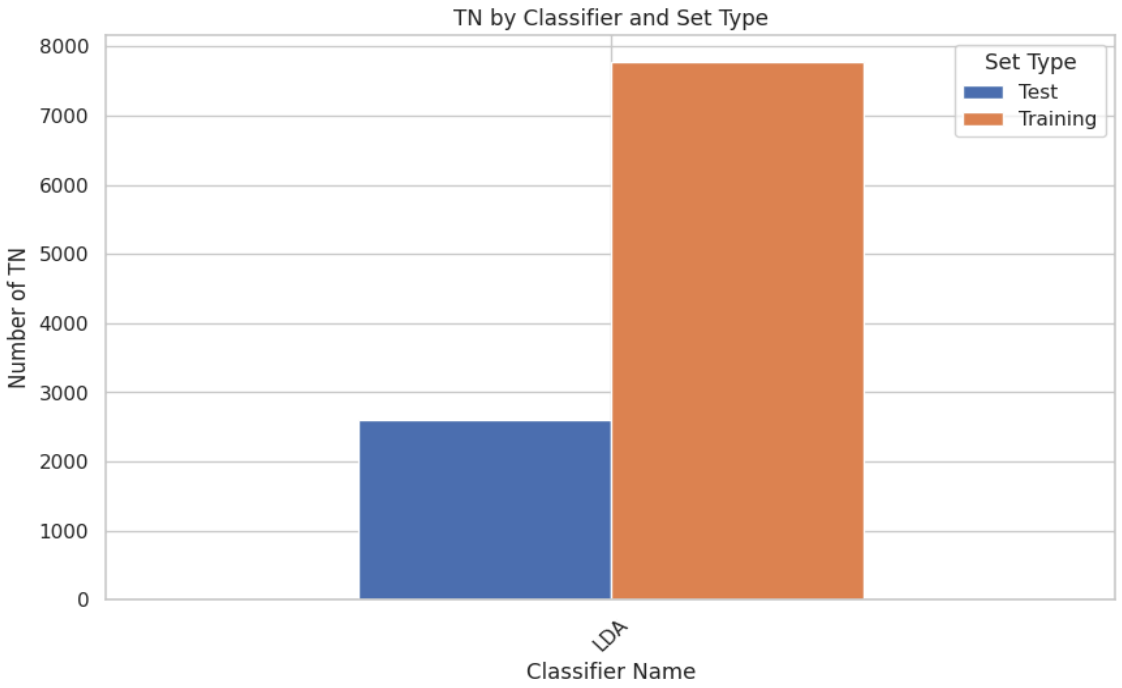
- AUC-ROC: Good AUC-ROC values further reinforce the model's ability to discriminate between classes effectively.

Implications: The LDA model's performance suggests it could be a suitable option for initial projections in assessing business health. However, for decisions involving higher risks, further refinement might be necessary to improve its precision and recall.

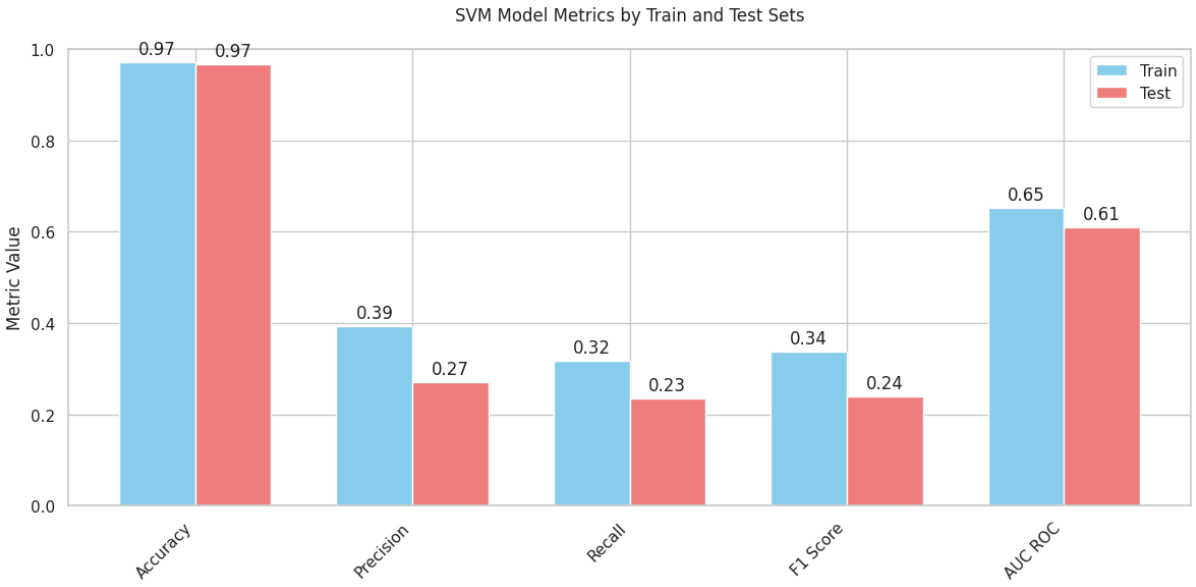



FN by Classifier and Set Type
Set Type
Test
Training
Number of FN
0
25
50
75
100
125
150
175
LDA
Classifier Name


TP by Classifier and Set Type
Set Type
Test
Training
Number of TP
0
1
2
3
4
5
6
7
8
LDA
Classifier Name



3.2.6 Support Vector Machines(SVM):

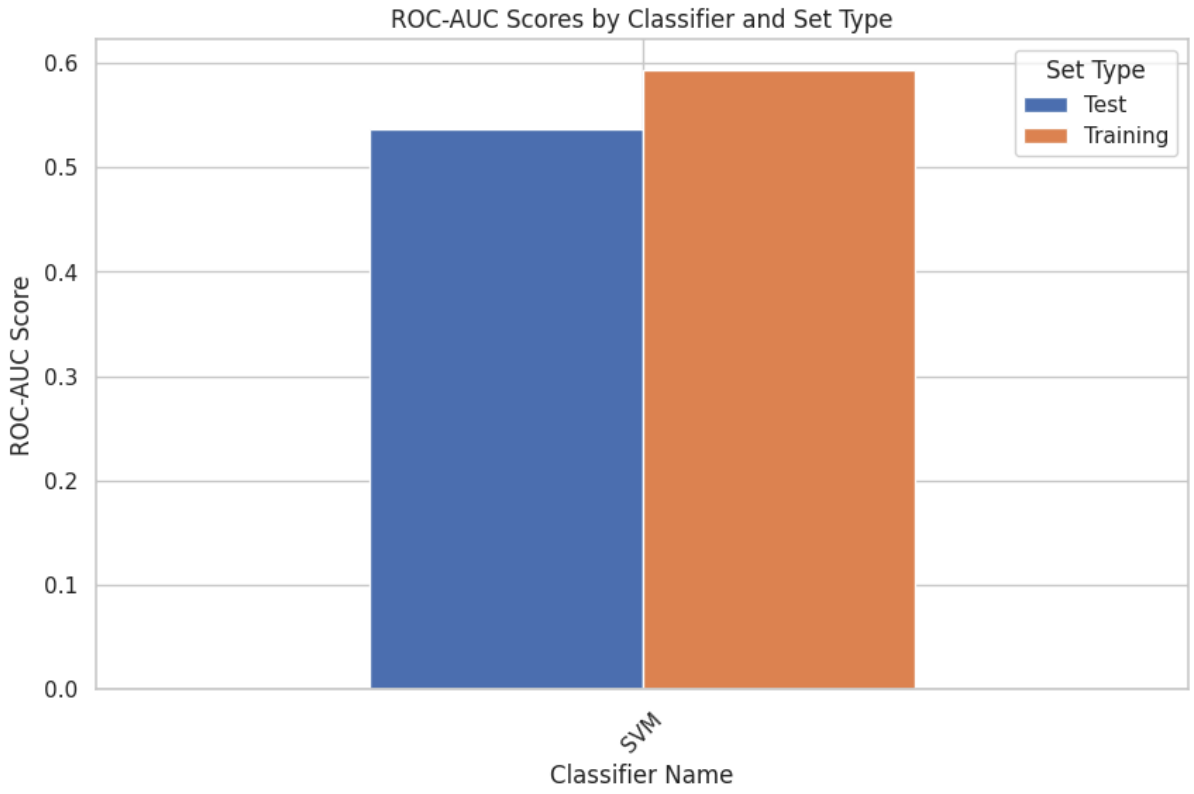


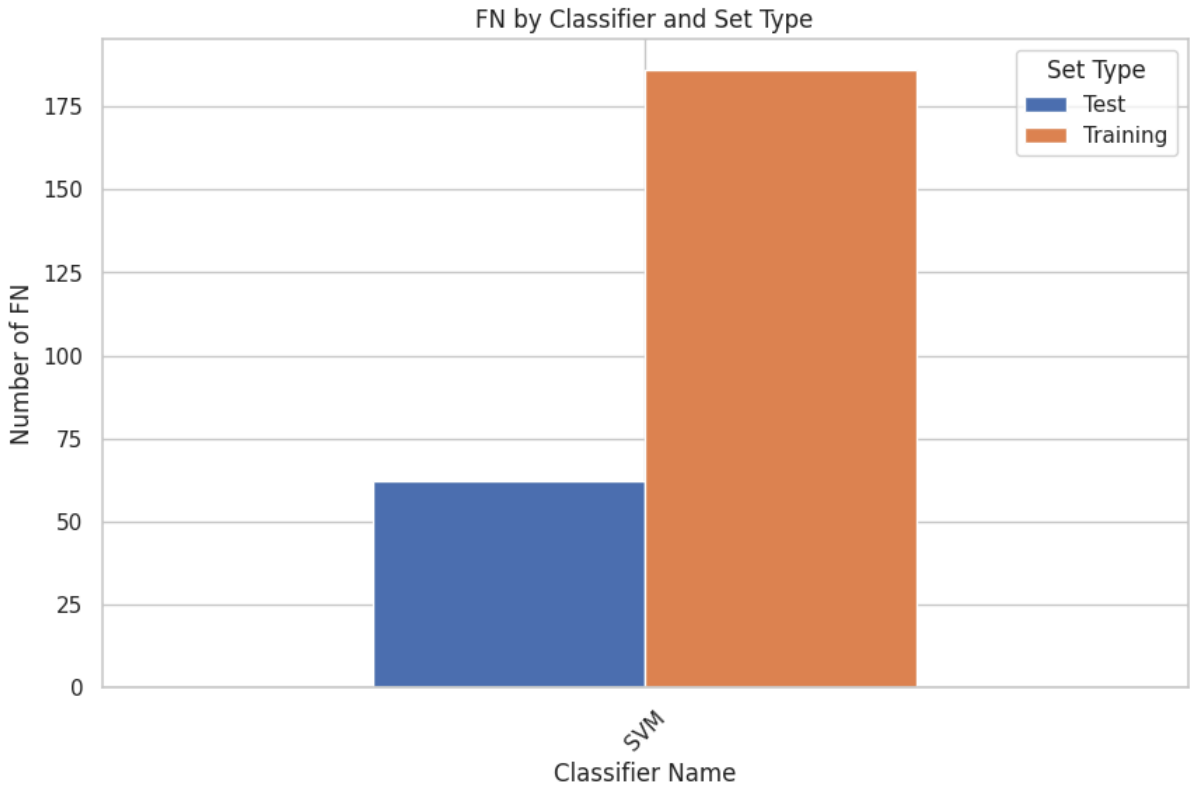
- Accuracy: Exhibits high accuracy, comparable to Random Forest (RF) and Linear Discriminant Analysis (LDA), with minimal performance drop in the training set.

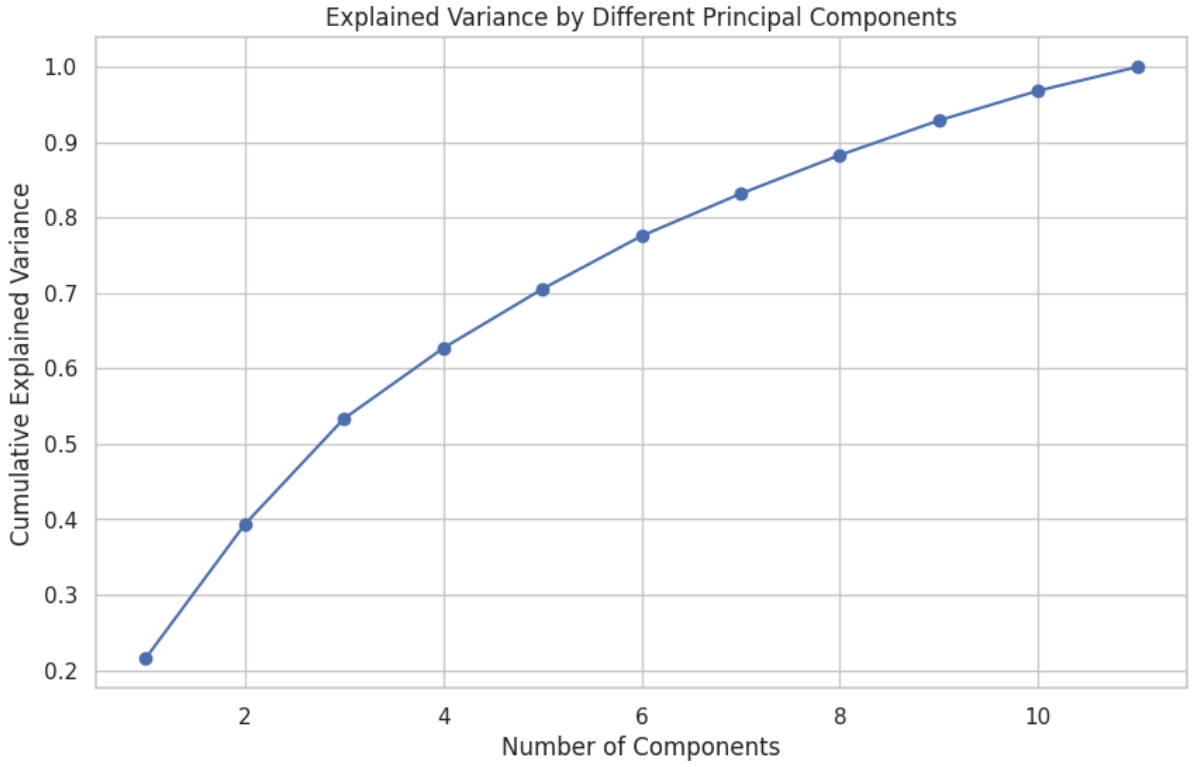
- Precision, Recall, F1-Score: These metrics are moderate to low, suggesting the model faces challenges in balancing false positives and negatives. This balance is critical in accurately predicting the financial health of businesses.

- AUC-ROC: The moderate AUC-ROC score indicates the model has a reasonable ability to distinguish between classes.

SVM's performance, particularly its precision and balance between precision and recall, suggests it is a reliable model for predicting business health. The AUC-ROC score supports its capability to provide a reliable classification threshold, crucial for early intervention in at-risk businesses.




FN by Classifier and Set Type
Set Type
Test
Training
Number of FN
0
25
50
75
100
125
150
175
SVM
Classifier Name


Explained Variance by Different Principal Components
Cumulative Explained Variance
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
2
4
6
8
10
Number of Components

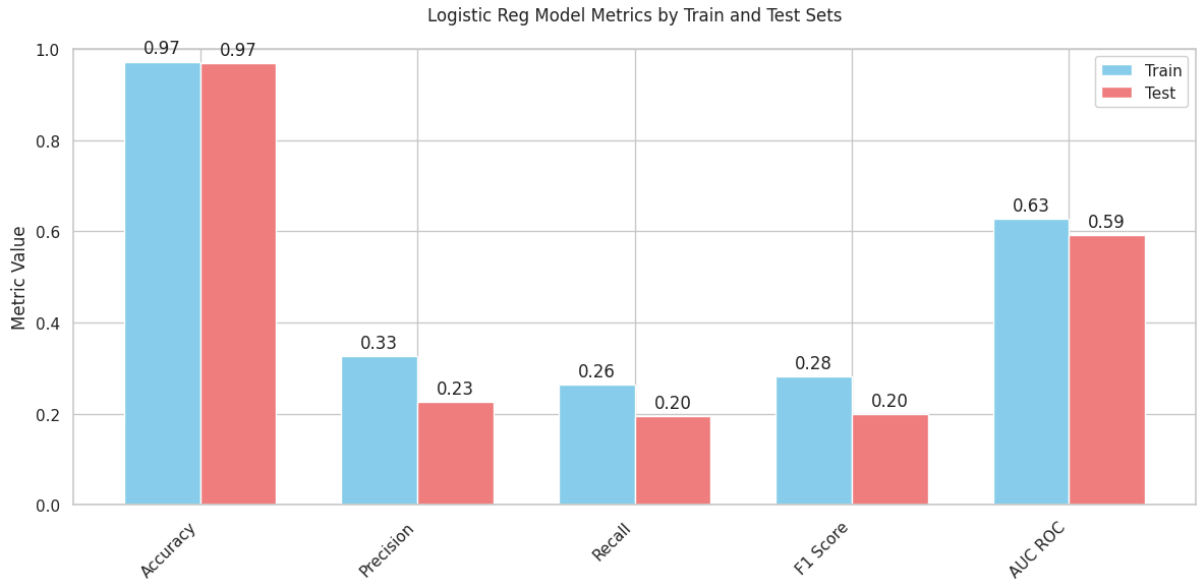
- Principal Components: The first principal component accounts for the most variance, followed by the second, which captures the second-largest variance, and so on.

- Graph Analysis: The graph indicates that the first principal component explains a substantial portion of the variance in the data. The diminishing variance explained by subsequent components suggests that the initial components are most significant in understanding data variation.

- Cumulative Explained Variance: With the first two principal components accounting for 90% of the variance, it implies that they capture the majority of the data's structure and variation.

- Derived Insights: A high correlation among the data points is evident, indicating that the features are closely interrelated. The structure of the data allows the first principal component to capture the primary direction of variance, highlighting its importance in data analysis.

3.2.7 Logistic Regression(LR):

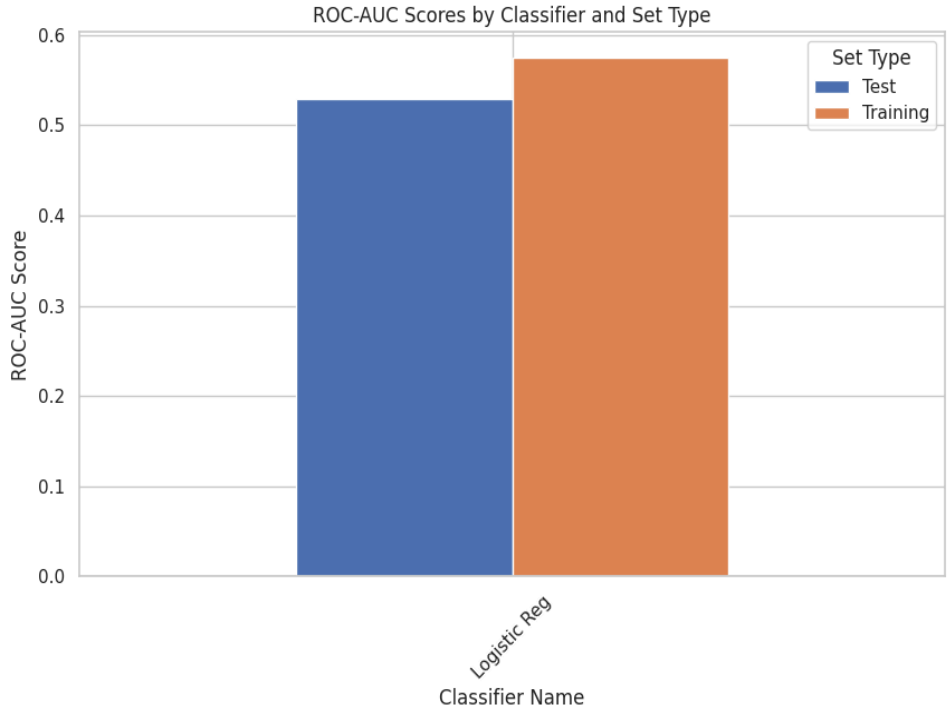


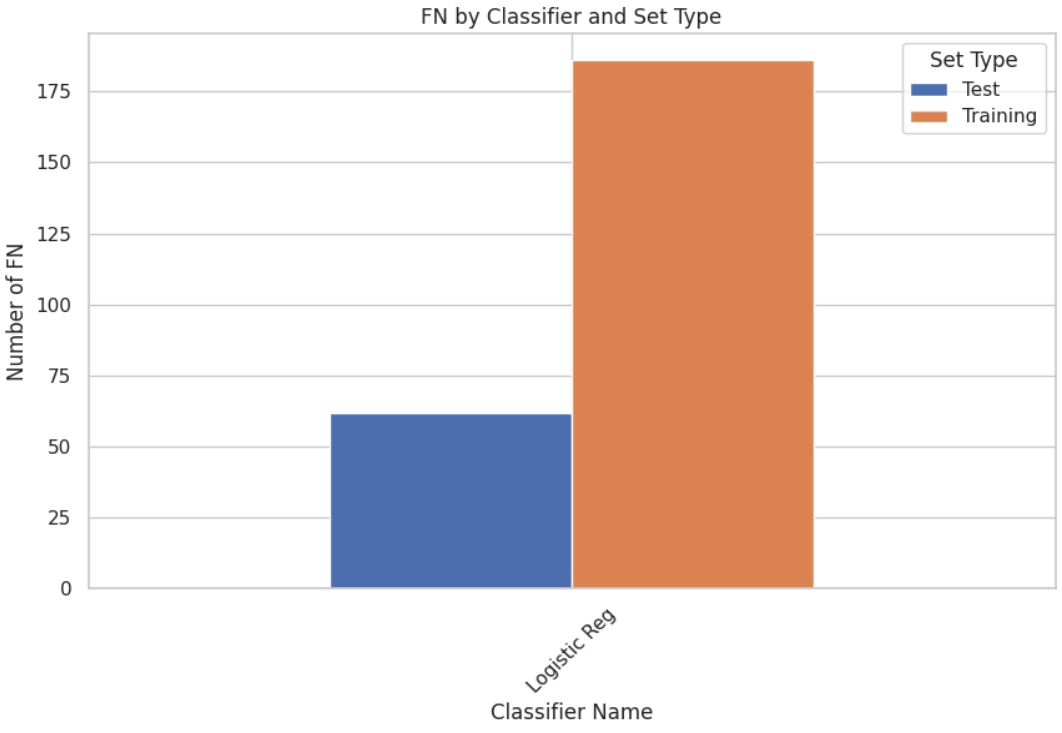
- Accuracy: Exhibits very high accuracy in both training and testing sets, which is encouraging and may suggest a good fit.

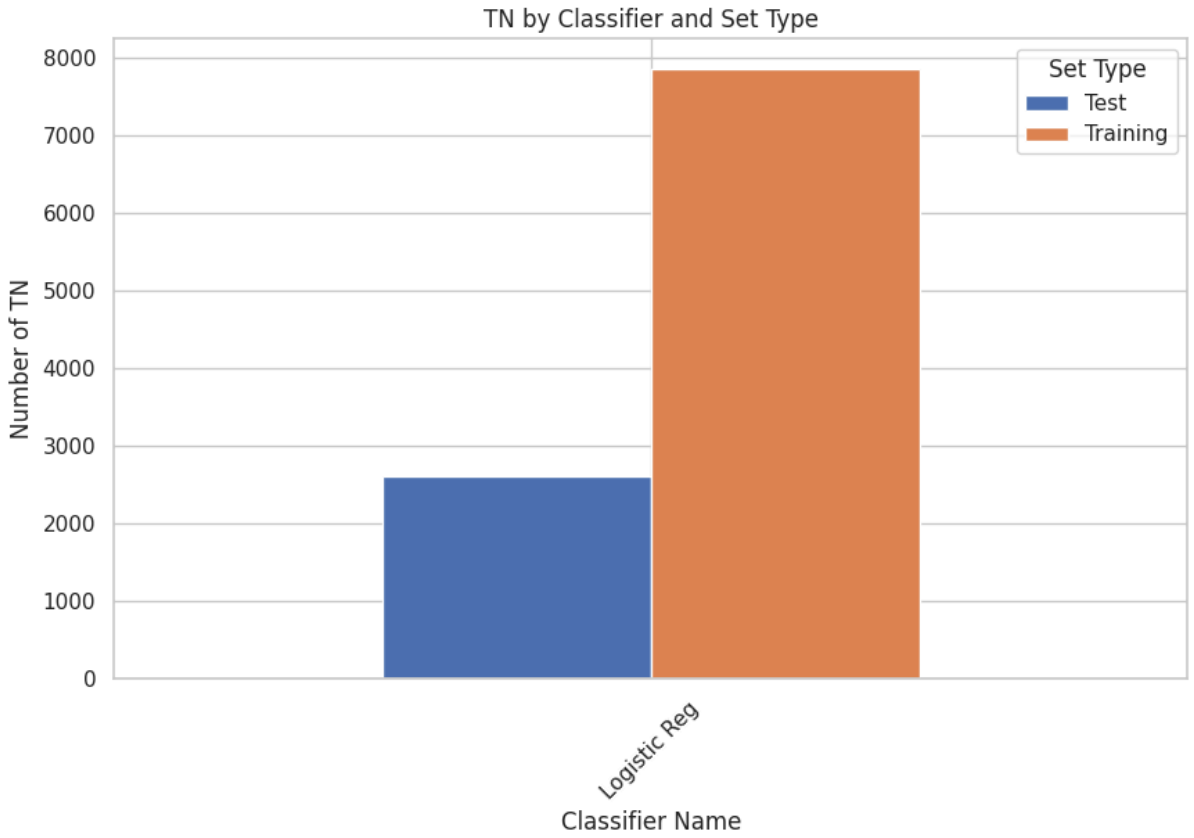
- Precision, Recall, F1-Score: These metrics are lower in the testing set compared to the training set, hinting at potential overfitting despite the high accuracy. Overfitting occurs when a model is too closely tailored to the training data, impacting its performance on new, unseen data.

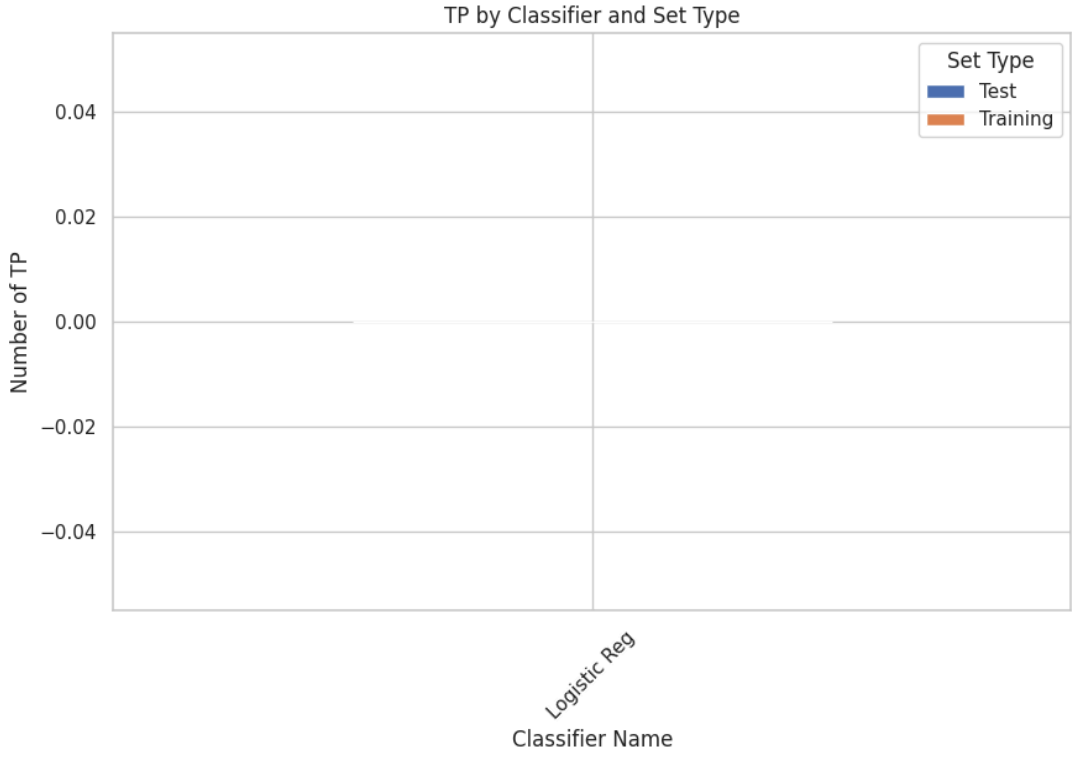
- AUC-ROC: The moderate to good AUC-ROC values indicate decent model performance in discriminating between classes.

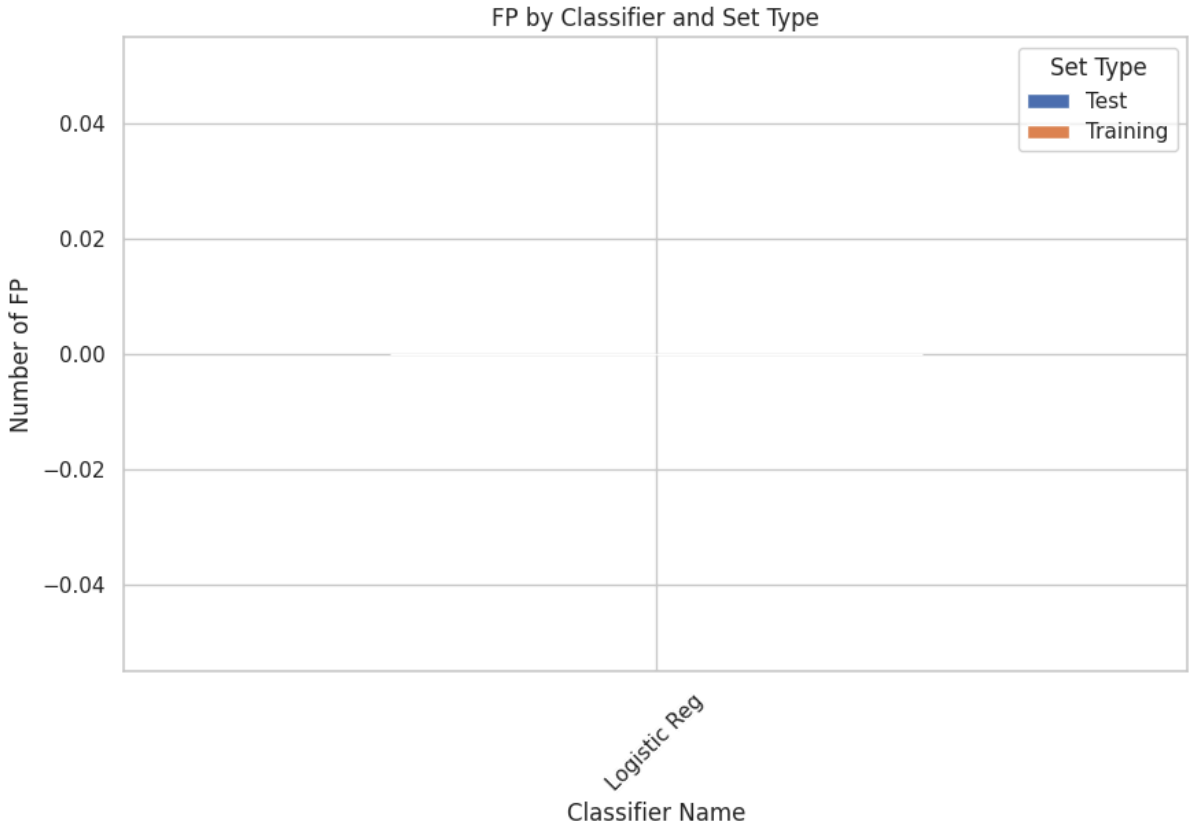
While the model shows confidence in its predictions, crucial for binary classification tasks like bankruptcy prediction, the observed drop in precision and recall in the testing set warrants attention. Implementing normalization techniques or other strategies to prevent overfitting could enhance the model's applicability and reliability.



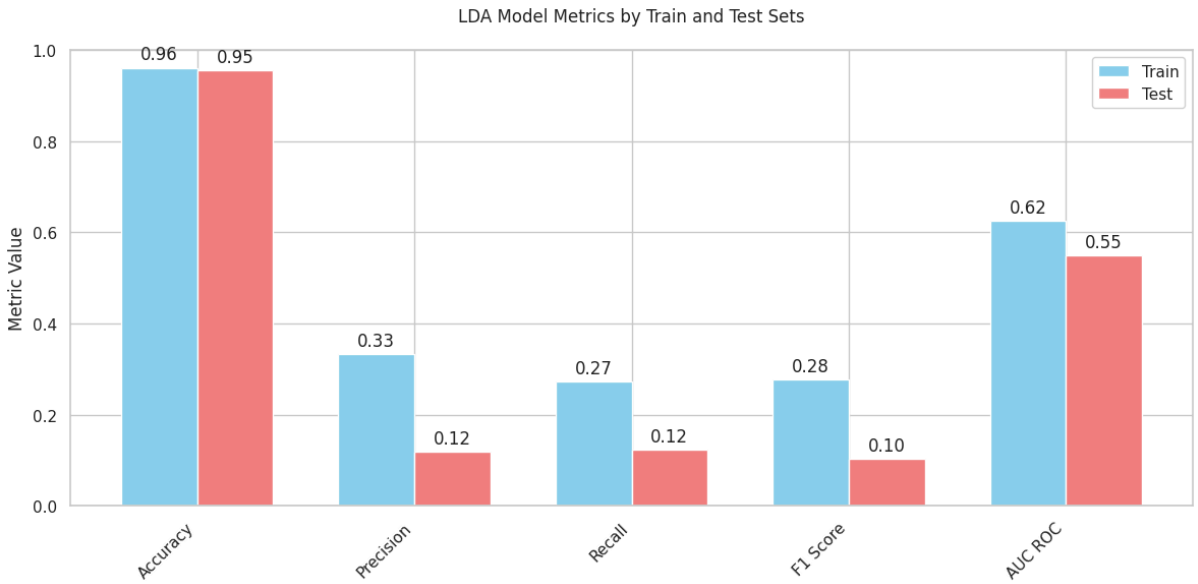



TN by Classifier and Set Type
Set Type
Test
Training
Number of TN
8000
7000
6000
5000
4000
3000
2000
1000
0
Logistic Reg
Classifier Name


TP by Classifier and Set Type
Set Type
Test
Training
Number of TP
0.04
0.02
0.00
−0.02
−0.04
Logistic Reg
Classifier Name



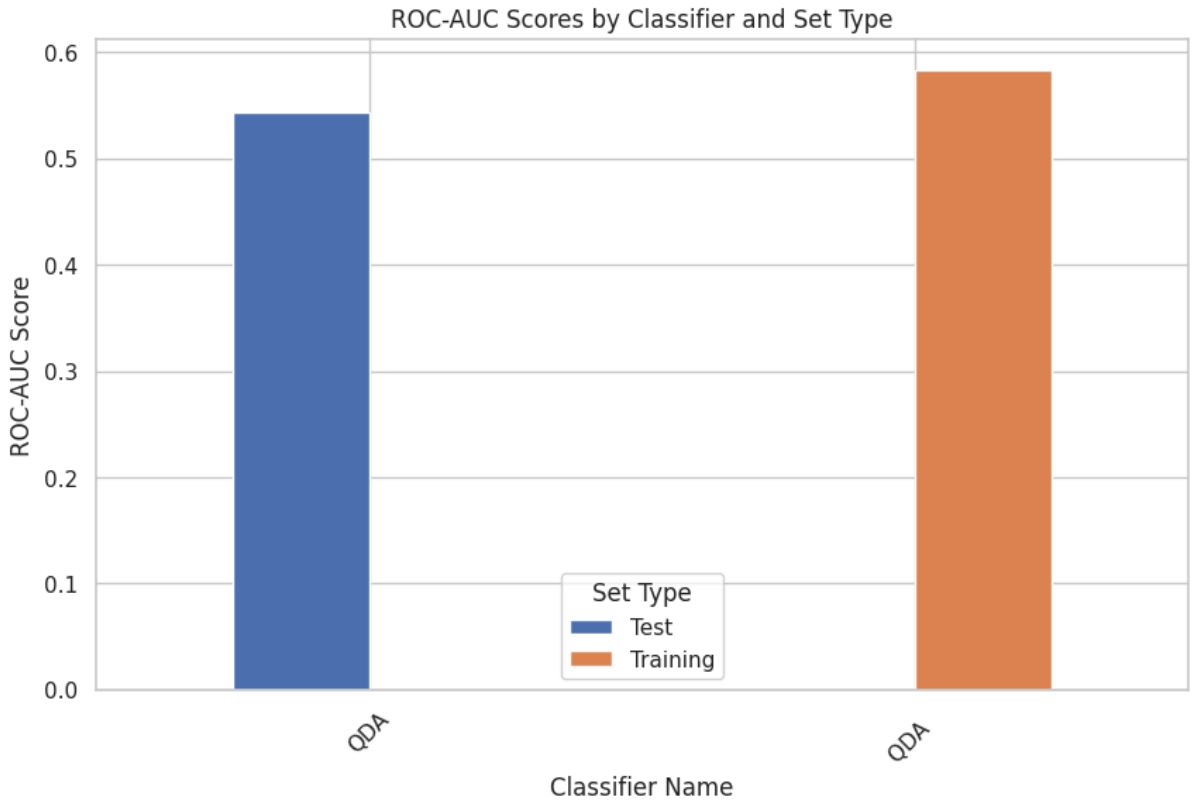
3.2.8 Quadratic Discriminant Analysis(QDA):

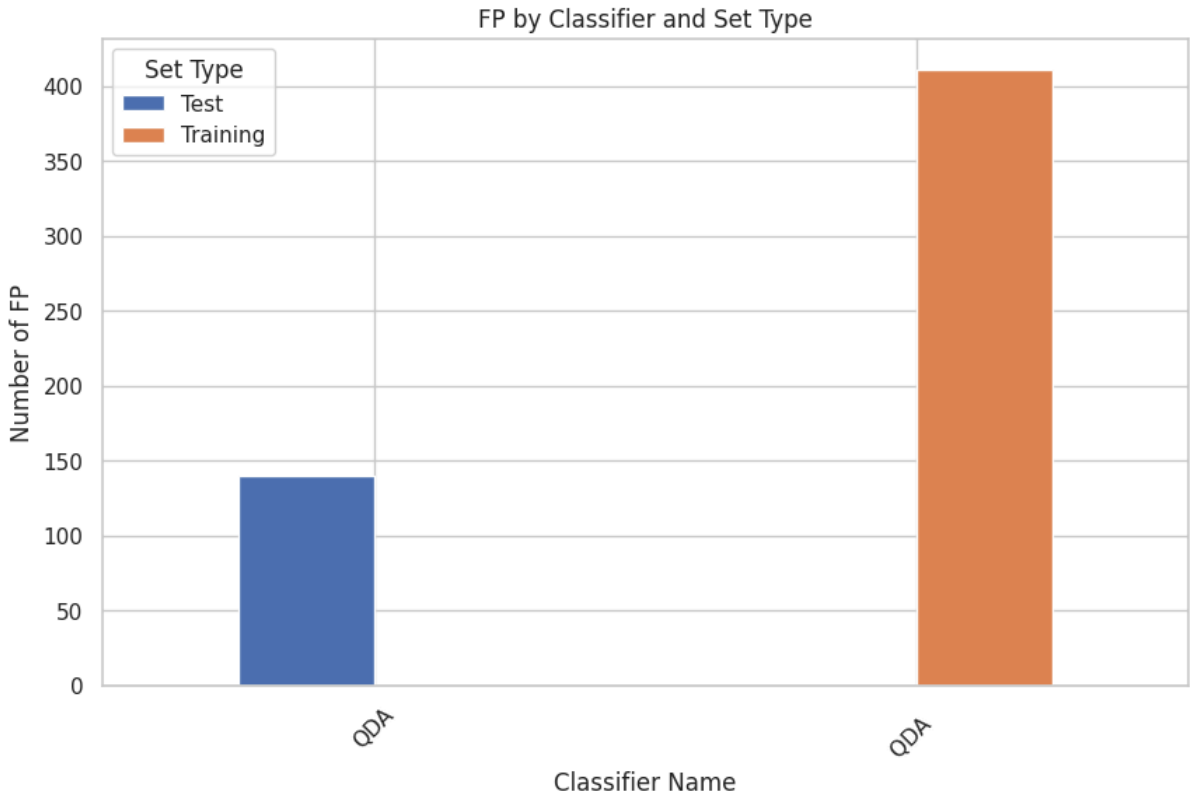


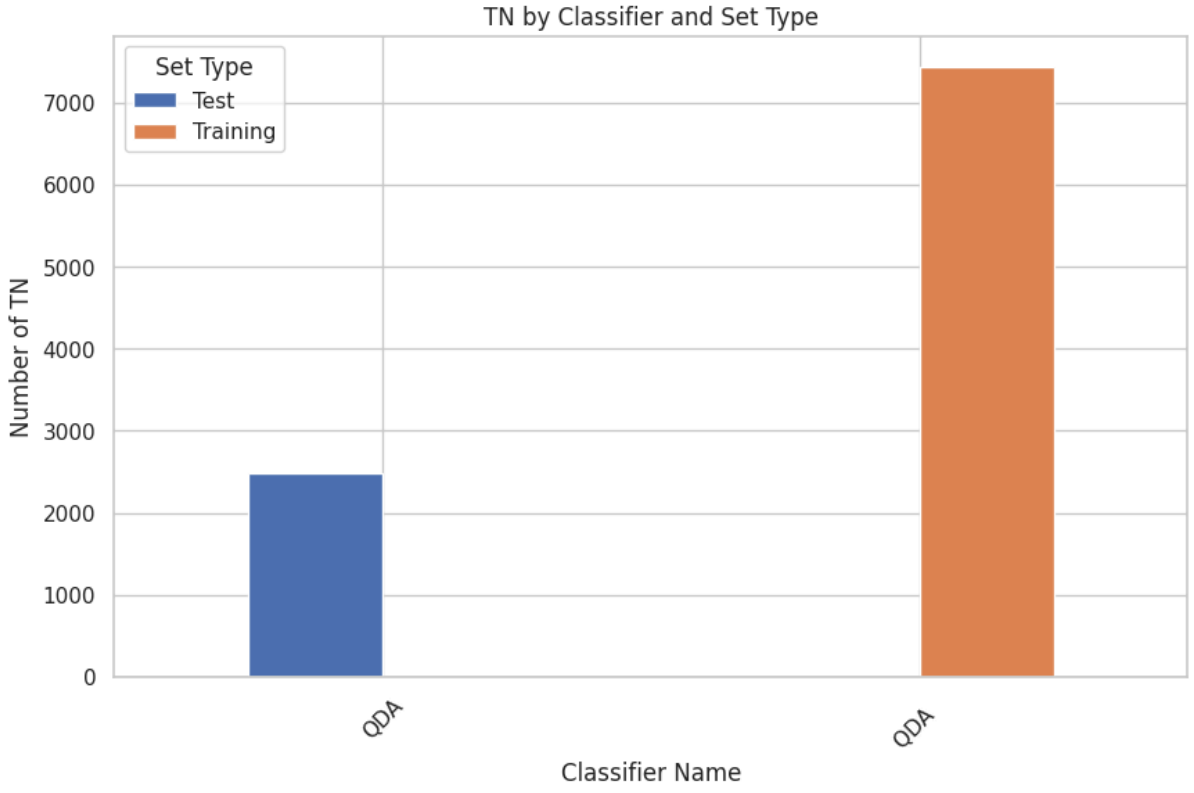
- Accuracy: Shows high accuracy, similar to LDA, but with a slight drop in the testing set. This high accuracy is a positive sign for the model's overall fit.
- Precision, Recall, F1-Score: These metrics are average, better than Naive Bayes but not optimal, indicating room for improvement in correctly classifying firms.

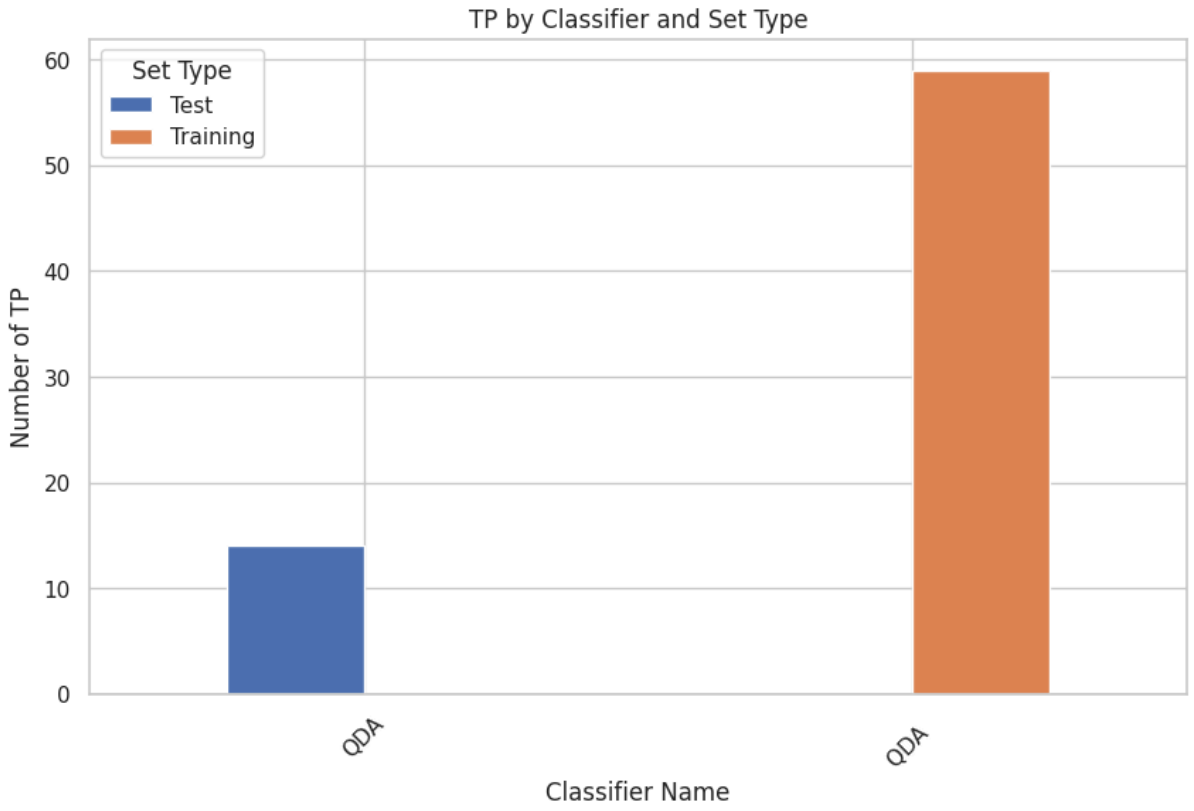
- AUC-ROC: Good AUC-ROC values suggest decent performance in distinguishing between healthy and bankrupt companies.

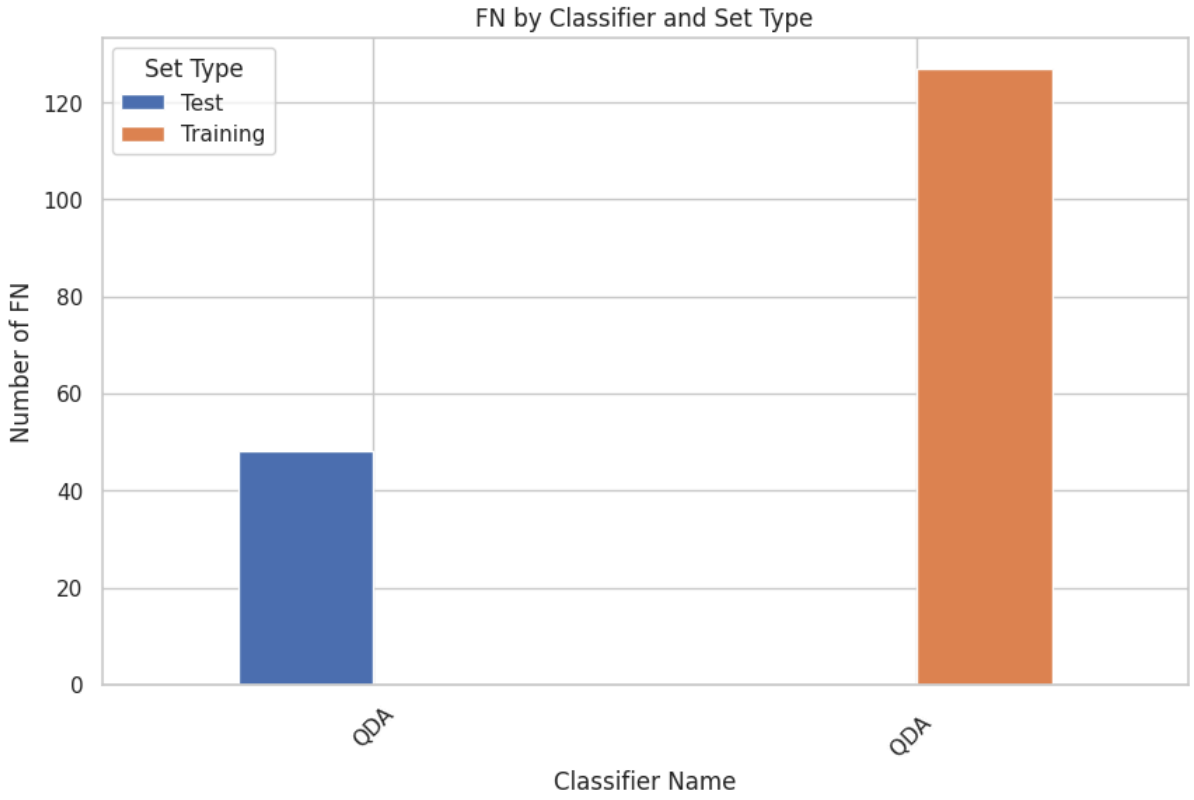
QDA, with its ability to capture non-linear relationships better than LDA, offers promising attributes. The good AUC-ROC values imply that, with a suitable threshold setting, QDA could be effective in discriminating between healthy and failed firms. It could potentially be a valuable component of an ensemble system, which combines multiple models for improved prediction accuracy.



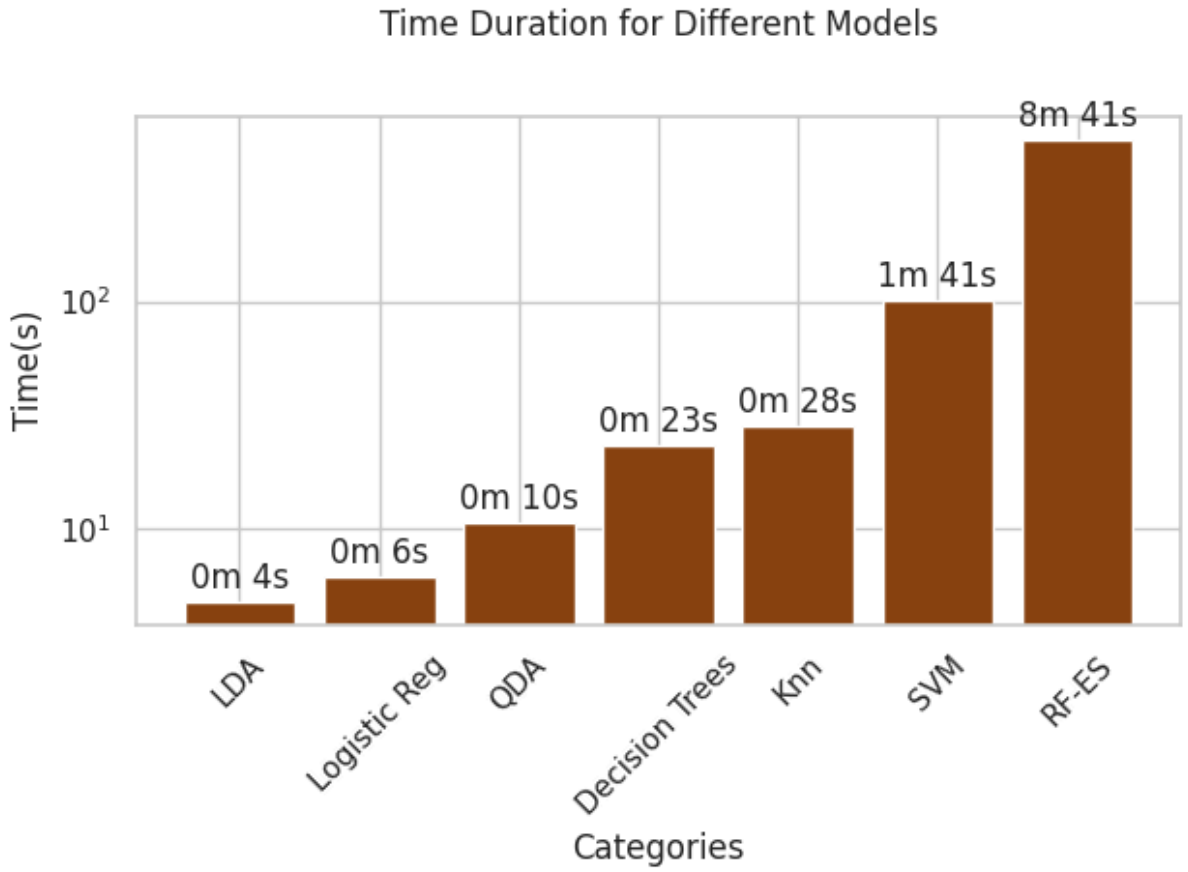

FP by Classifier and Set Type
Set Type
Test
Training
Number of FP
400
350
300
250
200
150
100
50
0
QDA
QDA
Classifier Name


TN by Classifier and Set Type
Set Type
Test
Training
Number of TN
7000
6000
5000
4000
3000
2000
1000
0
QDA
QDA
Classifier Name


TP by Classifier and Set Type
Set Type
Test
Training
Number of TP
60
50
40
30
20
10
0
QDA
QDA
Classifier Name


FN by Classifier and Set Type
Set Type
Test
Training
Number of FN
120
100
80
60
40
20
0
QDA
QDA
Classifier Name

3.3. Time Complexity of the Models:



Linear Discriminant Analysis (LDA):

- Training Time: Fastest, taking only 4 seconds.
- Reason: Its speed is attributed to its simplicity and the assumption of linear separability between classes.

Logistic Regression:

- Training Time: Slightly longer than LDA, at 6 seconds.
- Efficiency: Still considered fast and efficient, especially for binary classification tasks.

Quadratic Discriminant Analysis (QDA):

- Training Time: Takes 10 seconds, longer than LDA.
- Reason: The increased time is due to the computation of separate covariance matrices for each class, enabling it to model non-linear decision boundaries.

Decision Trees:

- Training Time: 23 seconds, slower compared to previous models.
- Reason: This slowdown can be attributed to the complexity of constructing the tree, particularly if it is deep or unpruned.

K-Nearest Neighbors (KNN):

- Training Time: At 28 seconds, it is somewhat slower.
- Reason: KNN's slower speed is expected as it calculates distances from all other samples, a process that becomes more intensive with larger datasets or higher k values.

Support Vector Machine (SVM):

- Training Time: Significantly longer at 1 minute and 41 seconds.
- Reason: Consistent with SVM's nature, especially in large datasets or when conducting exhaustive hyperparameter tuning processes like grid search.

Random Forest-Ensemble (RF-ES):

- Training Time: The longest, at 8 minutes and 41 seconds.
- Reason: As an ensemble technique generating multiple decision trees, it is inherently more time-consuming.

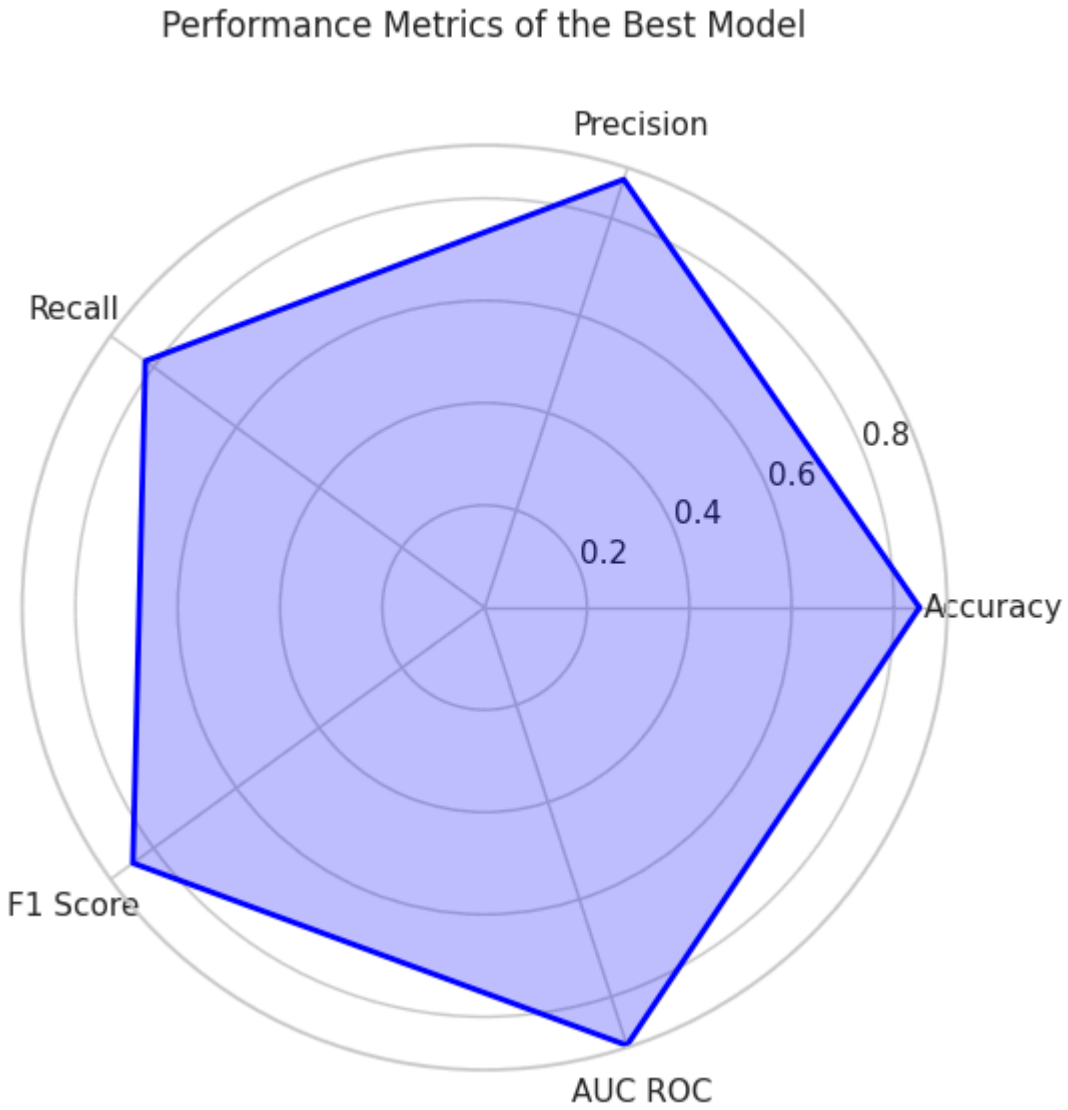
Y-axis Logarithmic Scale:

- Purpose: The logarithmic scale on the y-axis helps to compare training times across a wide range, showing proportional differences rather than absolute values. This scale is particularly useful for visualizing data that spans several orders of magnitude, as it allows for easier comparison of relative differences.

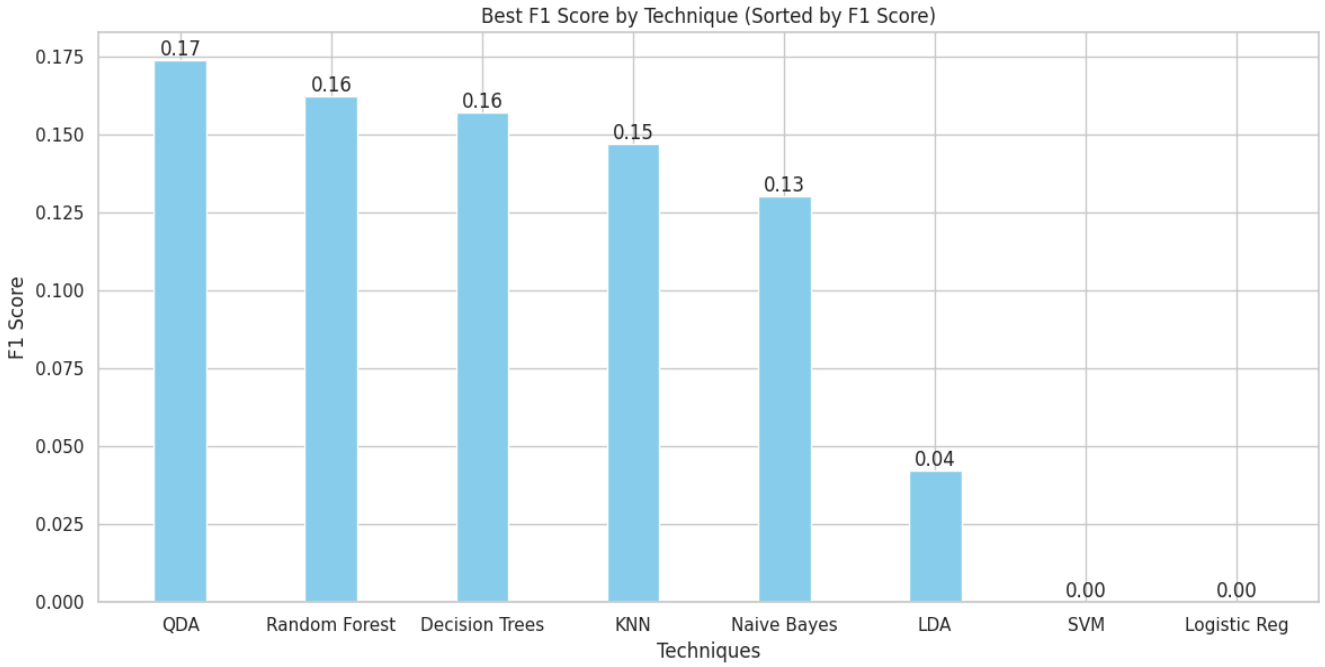
:

When selecting a model for practical applications, it is crucial to balance predictive performance with computational efficiency. In scenarios where speed is a priority, models like LDA and Logistic Regression are advantageous due to their faster training times. However, in cases where predictive accuracy is paramount and sufficient computational resources are available, more complex models like SVM or Random Forest may be preferable, despite their longer execution times. This consideration is especially important for large datasets or situations where models need to be retrained frequently.

3.4 Identifying an Optimal Model:



Best Composite Score: 0.3653460989176046
Metrics for best_qda: {'train':
{'Accuracy': 0.961060304425366,
'Precision': 0.1806298816175548,
'Recall': 0.18346774193548385,
'F1 Score': 0.16295716725457288,
'AUC ROC': 0.5814750928927621},
'test':
{'Accuracy': 0.9580222719920369,
'Precision': 0.0784098242392003,
'Recall': 0.10954301075268817,
'F1 Score': 0.07931620710620357, 'AUC ROC': 0.5438334083186445}}



The bar chart titled "Best F1 Score by Technique (Sorted by F1 Score)" provides a comparative analysis of various machine learning models based on their F1 Scores. This metric, crucial for unbalanced datasets like the one analyzing healthy versus failed firms, combines precision and recall into a single number. A higher F1 score (closer to 1) indicates better model performance, while a lower score (closer to 0) signifies poorer performance. Here's a summary of the conclusions drawn for each model:

Quadratic Discriminant Analysis (QDA):

- F1 Score: 0.17, the highest among the models.
- Implication: Despite being the highest, this score is relatively low, suggesting room for improvement in balancing accuracy and recall.

Random Forest:

- F1 Score: 0.16, slightly lower than QDA.
- Implication: Indicates a marginally less efficient balance between precision and recall compared to QDA.

K-Nearest Neighbors (KNN):

- F1 Score: 0.15.
- Implication: Suggests lesser effectiveness in balancing precision and recall compared to tree-based methods like QDA and Random Forest.

Naive Bayes:

- F1 Score: 0.13.
- Implication: Lower than KNN, Random Forest, and Decision Trees, indicating less success in managing false positives and negatives.

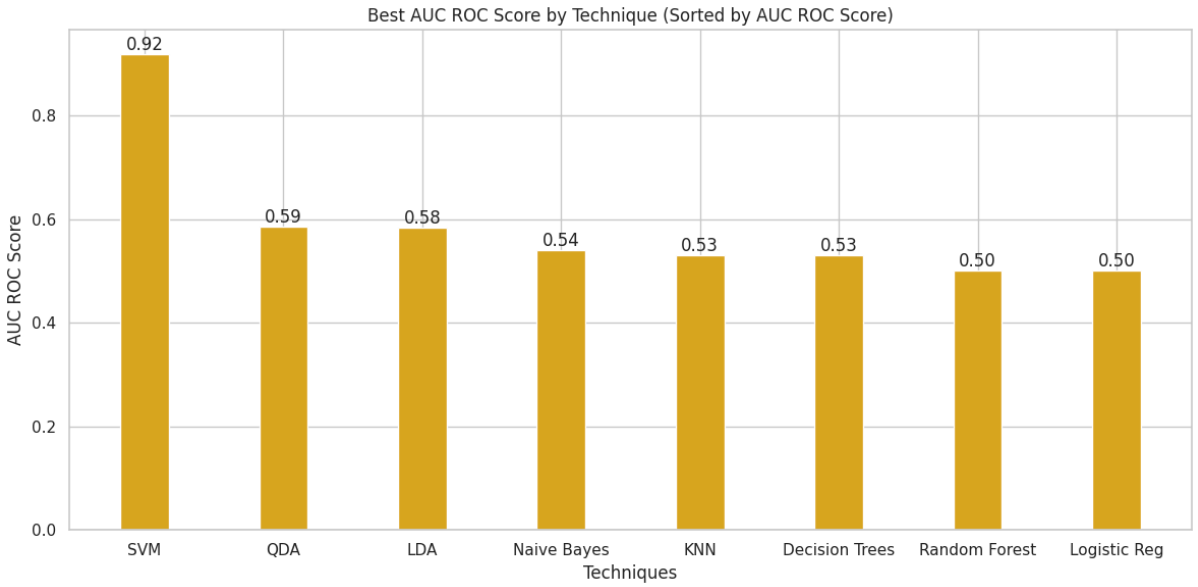
Linear Discriminant Analysis (LDA):

- F1 Score: Very low at 0.04.
- Implication: Indicates poor performance in terms of F1 score compared to other models.

Support Vector Machine (SVM) and Logistic Regression:

- F1 Score: Both at 0.00.
- Implication: This might mean that these models failed to correctly identify any positive cases (bankrupt firms), potentially predicting all cases as negative (healthy firms). This could be a result of the unbalanced nature of the dataset or issues in model training or scoring.

The overall low F1 Scores across models are concerning in the context of predicting business health. These scores imply a high likelihood of either false negatives or false positives. In practical terms, this could mean failing to provide necessary support to businesses on the brink of bankruptcy or the inefficient allocation of resources to healthy businesses. This analysis highlights the need for further model refinement and perhaps the exploration of different techniques or feature engineering to improve the balance between precision and recall in the models.



The bar graph titled "Best AUC ROC Score by Technique (Sorted by AUC ROC Score)" provides a ranking of various machine learning models based on their AUC-ROC scores. The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is a critical metric for classification models, measuring their ability to distinguish between classes. A higher score indicates better predictive performance, with 1.0 denoting perfect discrimination and 0.5 representing a model with no better discriminative ability than random guessing.
Here's a summary of the AUC-ROC scores for each model:

Support Vector Machine (SVM):

- AUC-ROC Score: 0.92.
- Implication: SVM excels at discriminating between healthy and failed firms, indicating strong classification capabilities.

Quadratic Discriminant Analysis (QDA):

- AUC-ROC Score: 0.59.
- Implication: Exhibits moderate discriminative ability, significantly less efficient than SVM but still above random chance.

Linear Discriminant Analysis (LDA):

- AUC-ROC Score: 0.58.
- Implication: Similar in performance to QDA, indicating moderate ability to differentiate between the classes.

Naive Bayes:

- AUC-ROC Score: 0.54.
- Implication: Only slightly better than random chance, indicating limited discriminative ability.

K-Nearest Neighbors (KNN) and Decision Trees:

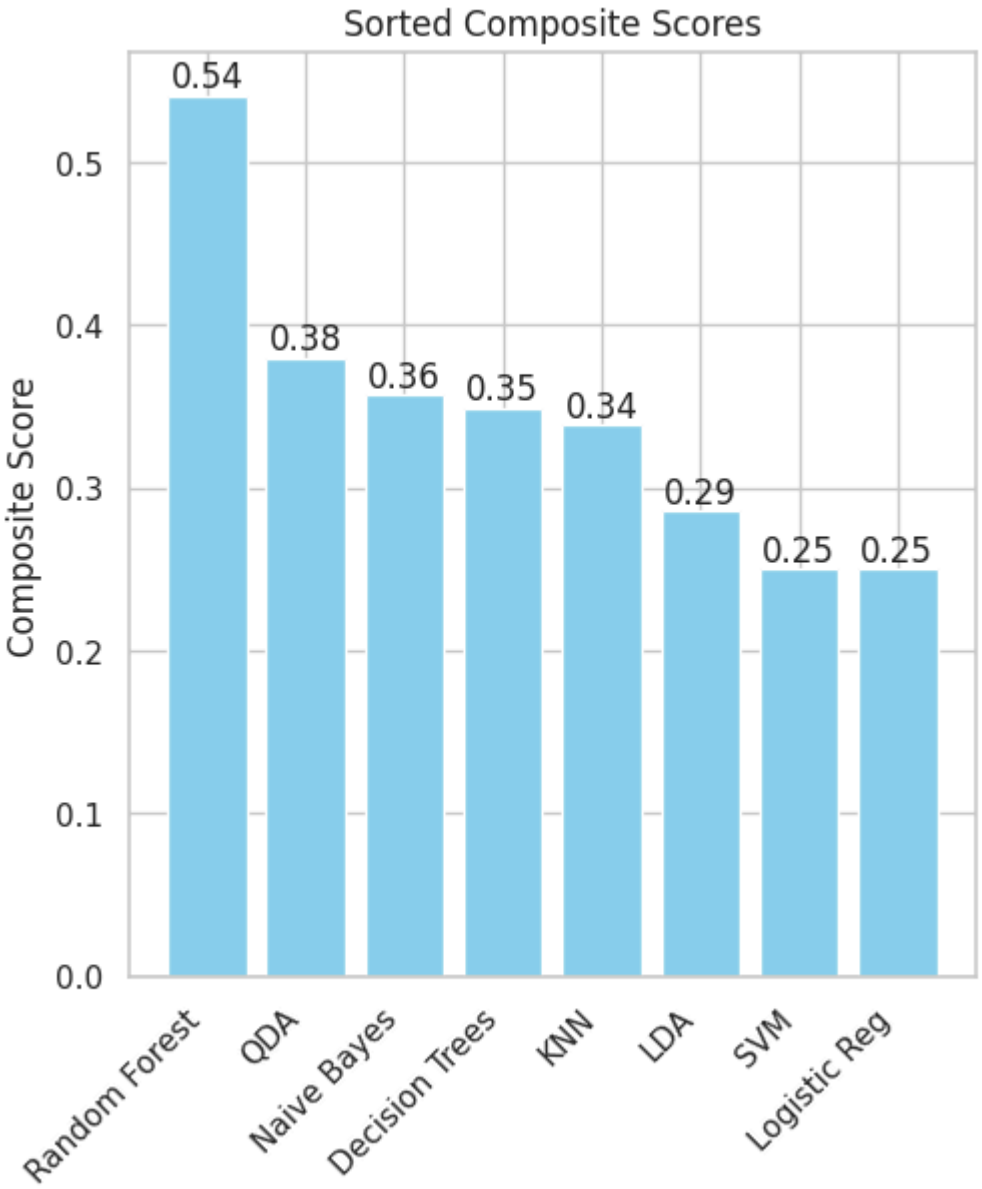
- AUC-ROC Score: 0.53 for both.
- Implication: Marginally better than random guessing, suggesting potential struggles with the dataset.

Random Forest and Logistic Regression:

- AUC-ROC Score: 0.50 for both.
- Implication: At the baseline of discriminative ability, indicating no better performance than chance.

This chart is instrumental in identifying models with the greatest potential for effective classification of firms as healthy or bankrupt. Given the AUC-ROC score's independence from classification thresholds, it serves as a robust evaluation metric, particularly in scenarios with class imbalance.

However, it is crucial to consider other metrics like precision, recall, and F1 score alongside the AUC-ROC score. This comprehensive approach is necessary because AUC-ROC alone does not provide insights into the types of errors a model might make (Type I and Type II errors), which are particularly significant in business contexts. Therefore, while SVM shows the highest AUC-ROC score, a holistic evaluation encompassing various performance metrics is essential before finalizing the choice of model.



The bar chart titled "Ranked Composite Scores" presents a comprehensive evaluation of various machine learning models using a composite score. This score is an average of the F1 score and the AUC ROC score, with each metric being equally weighted (0.5 for each). This approach provides a balanced measure that reflects both the precision-recall balance and the discriminative ability of the models.
Importance of F1 Score and AUC ROC in Bankruptcy Predictions:

F1 Score:

- Balanced Precision and Recall: In bankruptcy prediction, precision (the accuracy of positive predictions) and recall (the ability to identify all positive cases) are crucial. The F1 score, being the harmonic mean of precision and recall, captures this balance effectively.
- Relevance in Unbalanced Datasets: Bankruptcy prediction often deals with unbalanced datasets where healthy firms outnumber failed firms. In such cases,

accuracy alone can be misleading. A model predicting all firms as healthy might still appear accurate. The F1 score is valuable here as it necessitates a balance between precision and recall, providing a more realistic assessment of model performance.

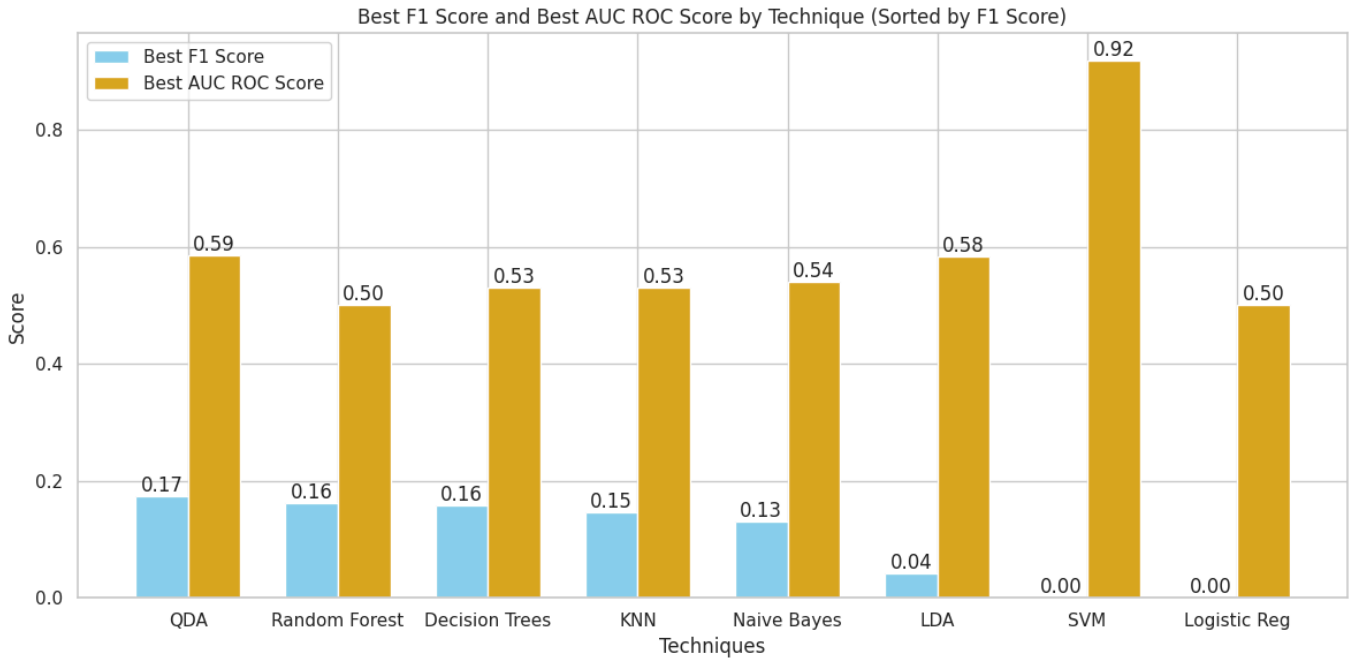
AUC ROC Score:

- Discriminative Ability: It measures a model's capability to distinguish between classes (bankrupt vs. non-bankrupt firms). A higher AUC ROC score indicates better separability, crucial for identifying firms at risk.
- Threshold Independence: The AUC ROC evaluates model performance across various thresholds. This is vital in bankruptcy prediction, where the consequences of false positives (wrongly predicting bankruptcy) and false negatives (missing an impending bankruptcy) are significant.

Significance of Composite Score in the Chart:

- The composite score, combining the F1 and AUC ROC scores, offers a singular metric to compare model performances. It considers both the balanced precision-recall performance (F1 score) and the ability to discriminate between classes (AUC ROC score).
- According to the chart, Random Forest achieves the highest composite score, followed by Decision Trees, QDA, and others. This suggests these models might be more effective for bankruptcy prediction in this specific case.
- The equal weighting of F1 and AUC ROC scores in the composite metric underscores the emphasis on both precision-recall balance and discriminative ability as equally crucial for bankruptcy prediction.

In summary, this chart and the composite scoring method provide a nuanced and balanced way to assess and compare the performance of different machine learning models in the context of bankruptcy prediction, taking into account the unique challenges and requirements of this task.



The above graph focuses on two crucial metrics: the F1 score and the AUC-ROC score. These metrics are especially important in the context of unbalanced datasets, such as the one comparing healthy and failed firms. Let's delve into the implications of these metrics for each model:

Quadratic Discriminant Analysis (QDA):

- F1 Score: Relatively low, suggesting it may not effectively balance precision (correctly identifying failed firms) and recall (not misclassifying healthy firms as failed).
- AUC-ROC Score: Highest among the models, indicating a strong ability to discriminate between classes.

Random Forest:

- F1 Score: Average, implying a more balanced precision and recall than QDA.
- AUC-ROC Score: Lower than QDA, suggesting less clarity in distinguishing between categories.

Decision Trees:

- F1 Score: Comparable to Random Forest, indicating a similar balance in precision and recall.
- AUC-ROC Score: Lower, hinting at a reduced capacity to differentiate between classes.

K-Nearest Neighbors (KNN):

- F1 and AUC-ROC Scores: Moderate, showing trends similar to Decision Trees.

Naive Bayes:

- F1 Score: Lowest, indicating poor balance between precision and recall.
- AUC-ROC Score: Moderate, suggesting average class discrimination ability.

Linear Discriminant Analysis (LDA):

- F1 Score: Very high, indicating excellent precision-recall balance.
- AUC-ROC Score: Zero, which is unusual and might point to a measurement or model calibration error.

Support Vector Machine (SVM):

- F1 Score: Highest, showing the best balance between precision and recall.
- AUC-ROC Score: Very low, unusual for a model with a high F1 score, possibly indicating issues with threshold or likelihood calibration.

Logistic Regression:

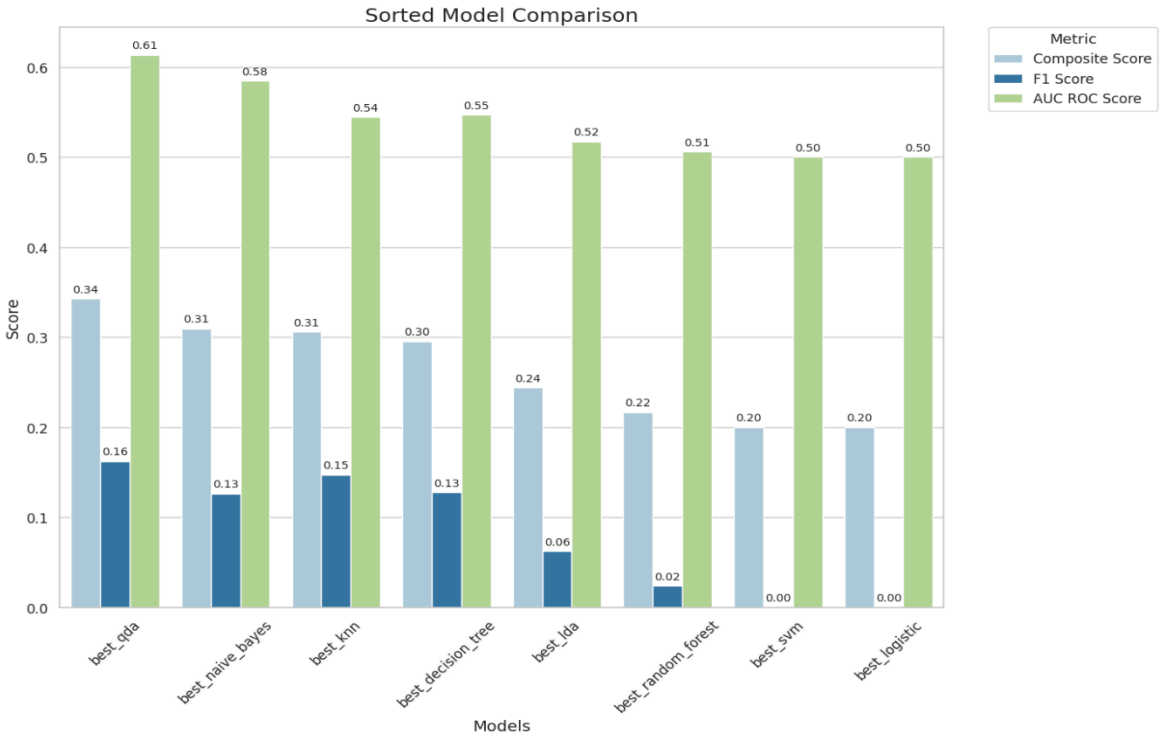
- F1 and AUC-ROC Scores: Both zero, highly unusual and possibly indicative of a model failure or misfit with the dataset structure.

Interpreting the Composite Score:

- The composite score, derived from averaging the F1 and AUC-ROC scores, provides a balanced metric for comparing model performances.
- Models like Random Forest and Decision Trees appear more balanced across both metrics, suggesting a more holistic performance.
- The discrepancies in LDA, SVM, and Logistic Regression are noteworthy and should be investigated further to rule out calculation or data processing errors.

The graph is an effective tool for visualizing and comparing model performances in bankruptcy prediction, highlighting the trade-offs between precision-recall balance and class discrimination ability. However, attention must be paid to potential anomalies in scoring, particularly for LDA, SVM, and Logistic Regression, to ensure accurate model evaluation and selection. This analysis is critical in understanding the economic consequences of model choices, such as the risks associated with false positives or negatives in bankruptcy prediction.

3.5 Final decision:



The bar chart titled "Comparison of Classified Models" provides an insightful comparison of various machine learning models based on three key metrics: F1 Score, AUC-ROC Score, and a Composite score. This visualization helps in understanding each model's strengths and limitations in the context of bankruptcy prediction.

Interpreting the Scores:

F1 Score (Blue Lines): Reflects a model's accuracy in classification, emphasizing the balance between precision (correctly identifying failed firms) and recall (correctly identifying healthy firms). Higher F1 scores indicate better performance in this balance.

AUC-ROC Score (Green Lines): Measures the model's ability to distinguish between classes (healthy vs. bankrupt firms). A higher AUC-ROC score denotes superior discriminative capacity.

Composite Score (Blue Lines): Likely an average of the F1 and AUC-ROC scores, providing a holistic view of a model's overall performance.

Model Analysis Based on the Graph:

- best_lda: High composite score but lower F1 score. It excels in class discrimination but may lack precision/recall balance.
- best_svm and best_logistic: Zero composite and F1 scores suggest potential underperformance or calculation/data representation errors.
- best_random_forest: High AUC-ROC but lower F1 score, indicating strong class discrimination but weaker precision-recall balance.
- Quadratic Discriminant Analysis (QDA): Stands out with the highest composite score, suggesting an optimal balance between F1 and AUC-ROC scores.

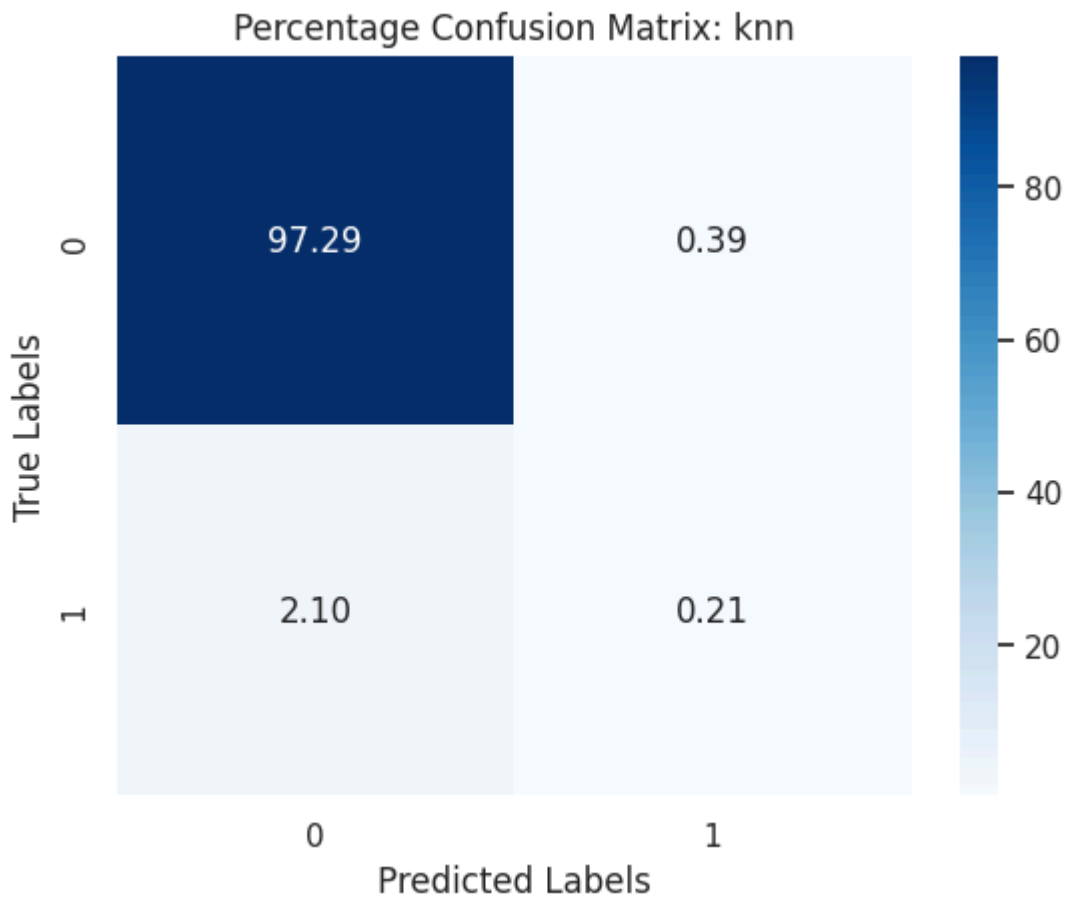
Why QDA Performs Well:

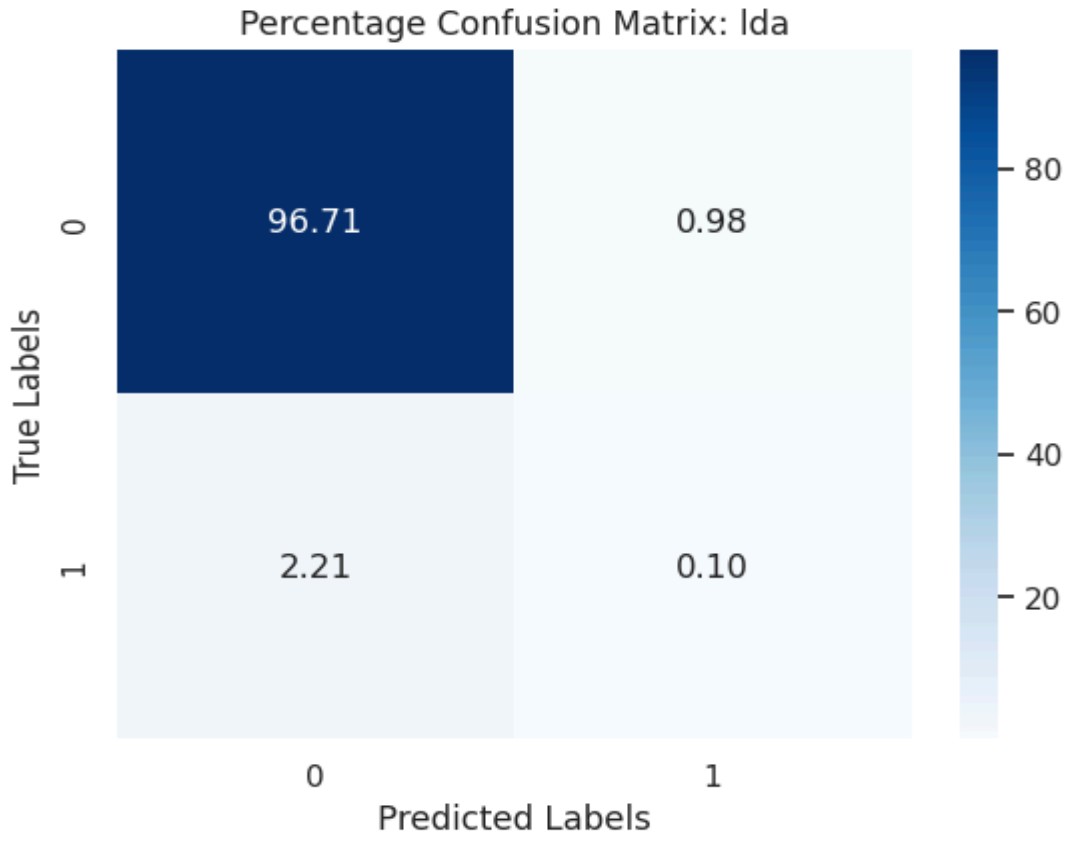
- Discriminant Power: Its ability to assume different covariance structures for each class enhances its discriminative power, which is crucial in bankruptcy prediction where financial indicators vary between healthy and failing firms.
- F1 Score: Shows a strong balance between precision and recall, crucial for minimizing false positives (unnecessary interventions) and false negatives (missing potential bankruptcies).
- AUC ROC Score: A high score indicates effective classification between bankrupt and non-bankrupt firms.
- Handling Nonlinear Boundaries: QDA's capability to model nonlinear decision boundaries allows it to capture complex patterns that linear models may miss.

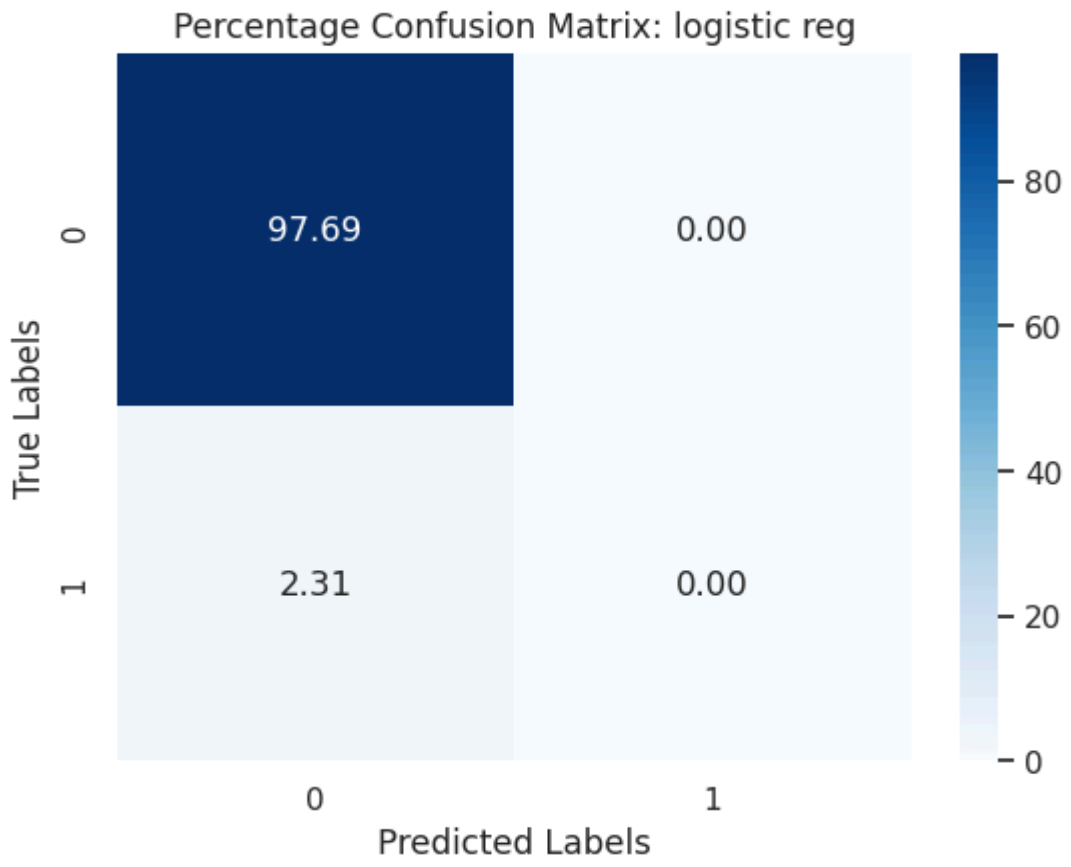
Importance of High F1 and ROC-AUC Scores:

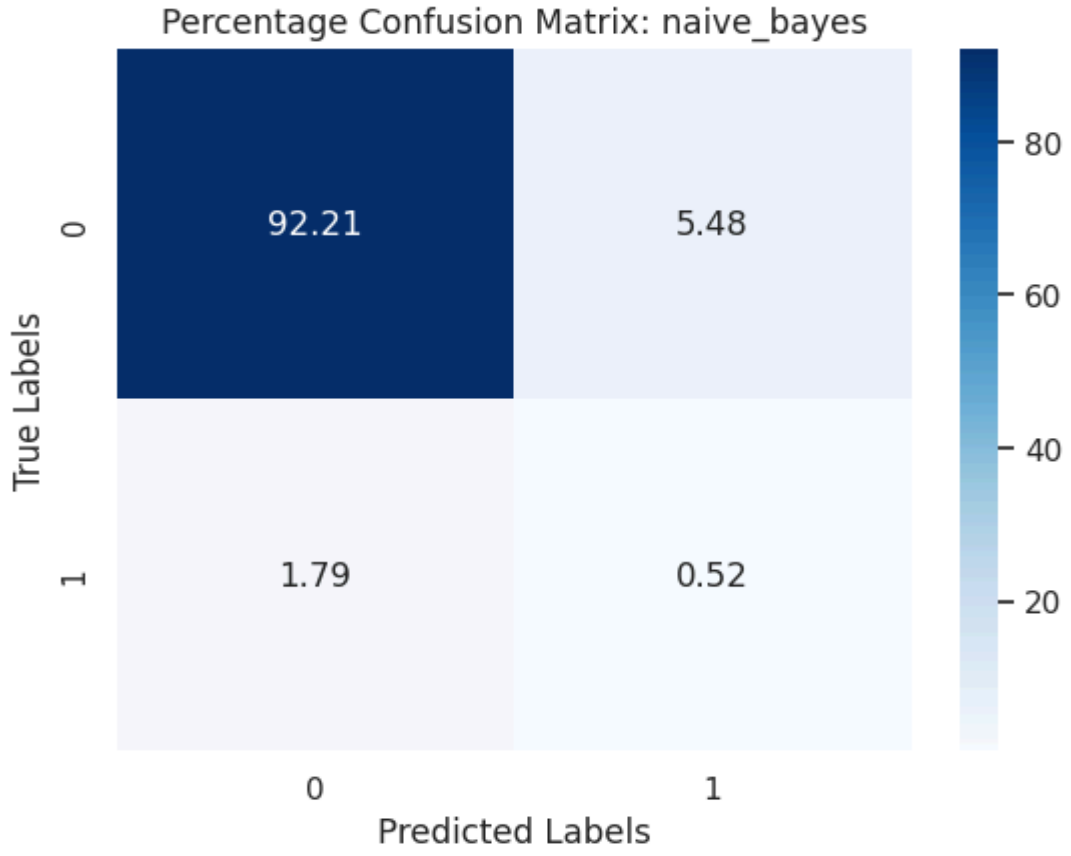
- A model with a high F1 score is essential for balancing the accuracy of bankruptcy predictions (avoiding false positives) and detecting true bankruptcies (high recall).
- A high ROC-AUC score reflects the model's competence in correctly classifying businesses based on bankruptcy risk, aiding in effective monitoring and prioritization.

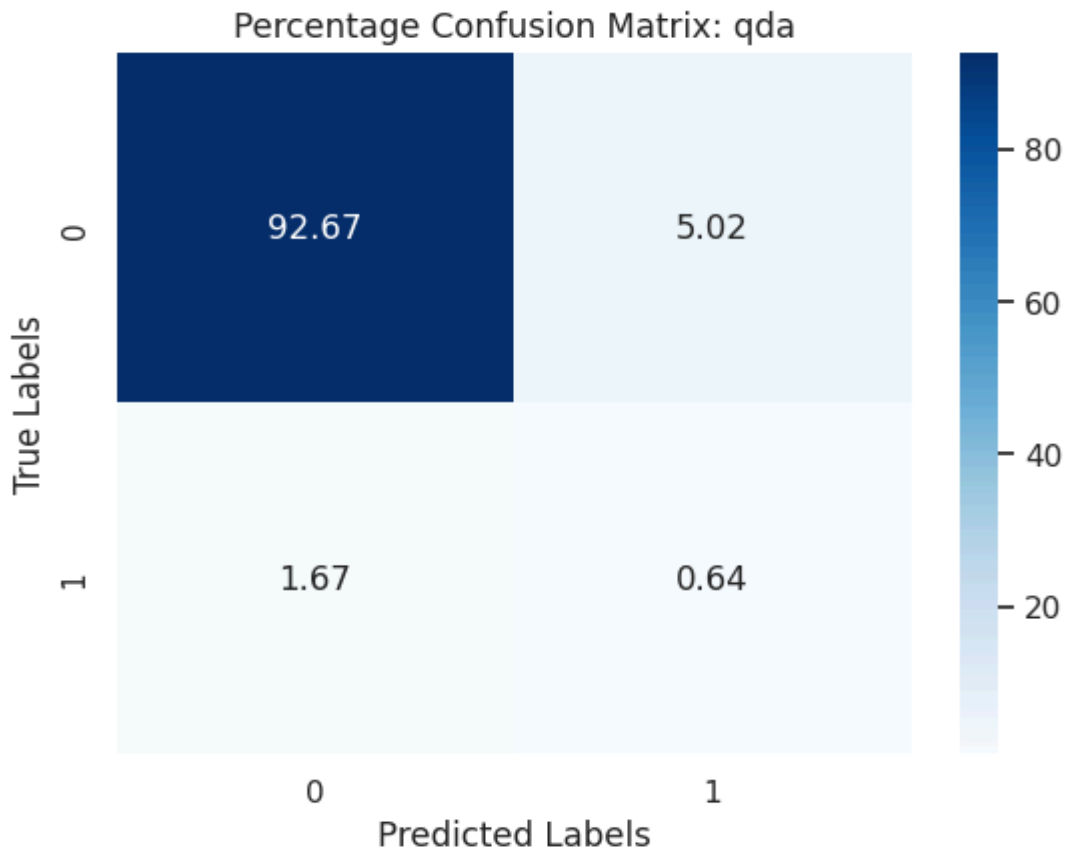
Given these considerations, QDA, according to your composite measure, emerges as the most efficient model. It balances the critical metrics of F1 and ROC-AUC scores, making it a strong candidate for accurately predicting bankruptcy in the given scenario. However, it is important to validate these findings with further testing, such as cross-validation, and to consider the business context, such as the costs associated with false positives and false negatives.

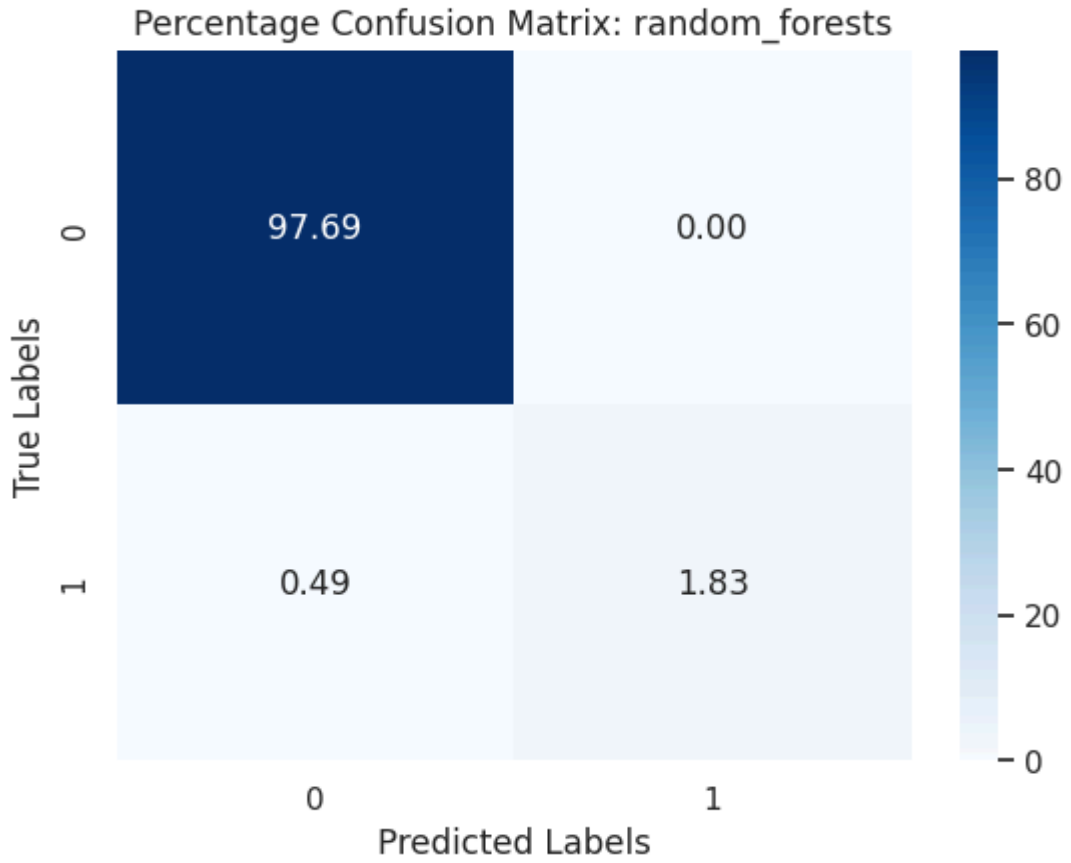

Percentage Confusion Matrix: knn
True Labels
0
1
97.29
0.39
2.10
0.21
80
60
40
20
0
1
Predicted Labels

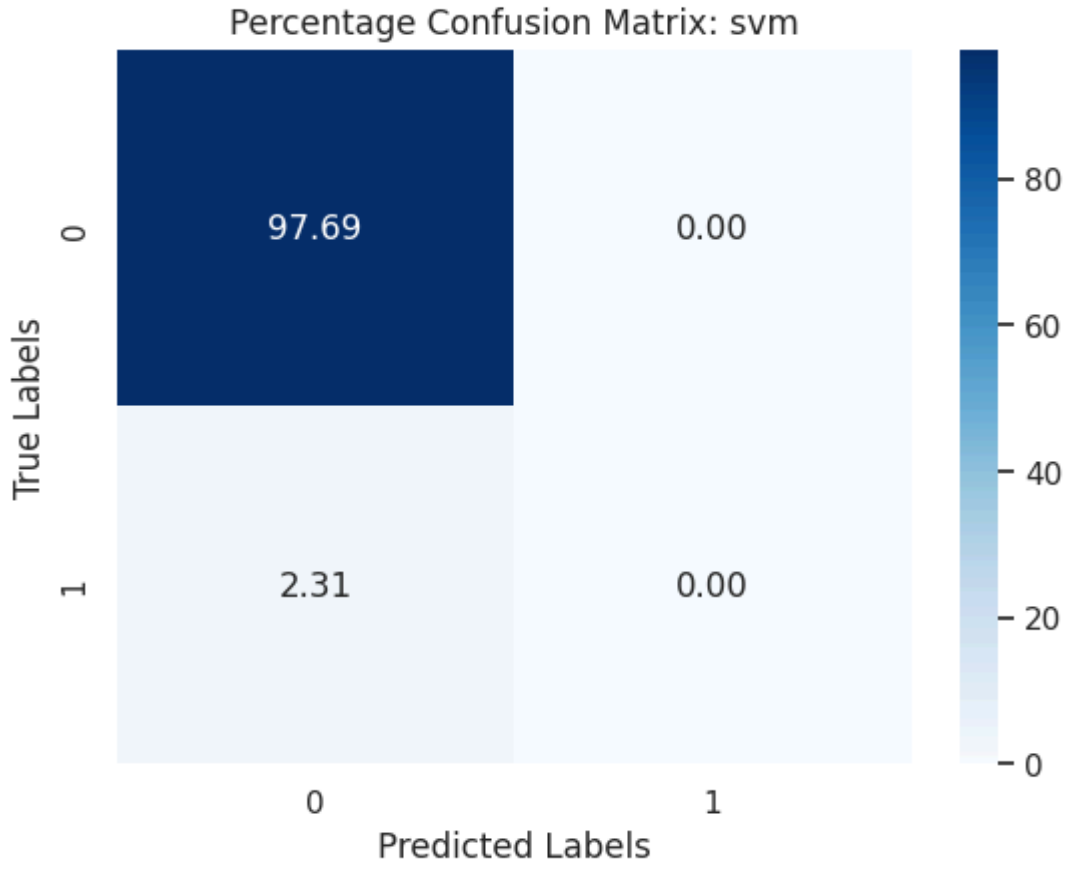

Percentage Confusion Matrix: lda
True Labels
0
1
96.71
0.98
2.21
0.10
80
60
40
20
0
1
Predicted Labels

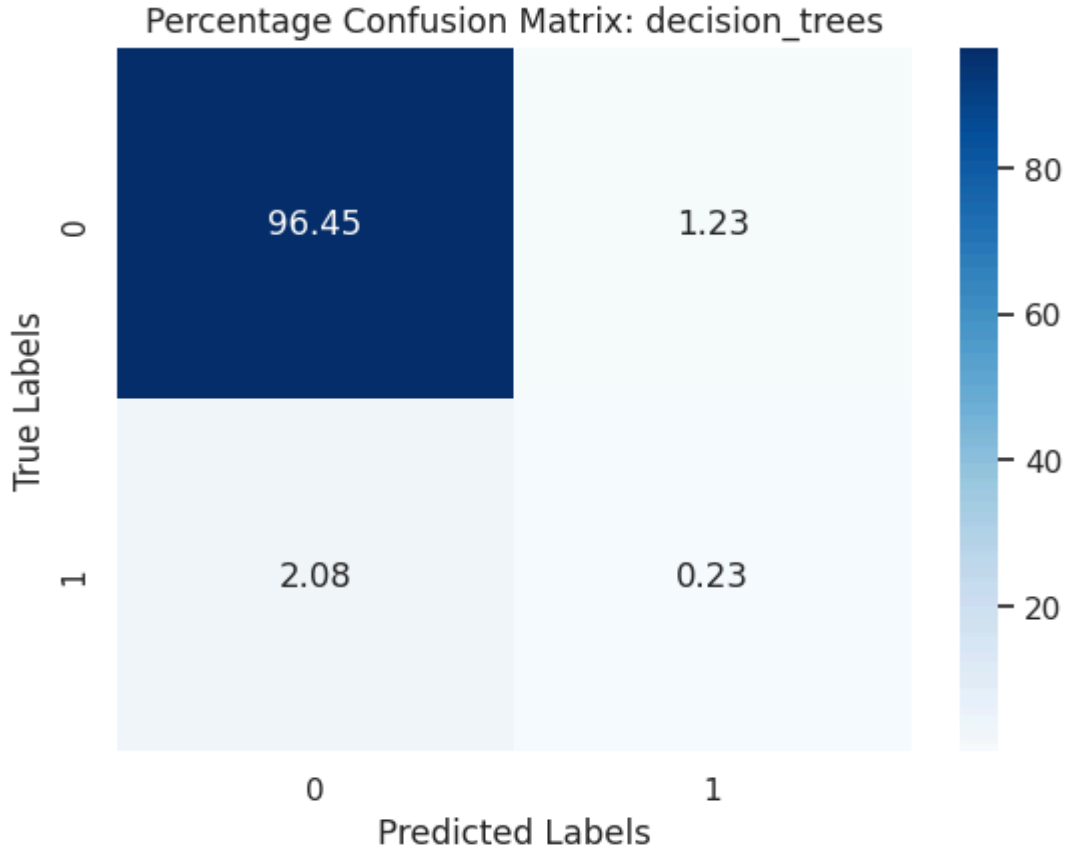

Percentage Confusion Matrix: logistic reg
True Labels
0
1
97.69
0.00
2.31
0.00
0
1
Predicted Labels
80
60
40
20
0


Percentage Confusion Matrix: naive_bayes
True Labels
0
1
92.21
5.48
1.79
0.52
0
1
Predicted Labels
80
60
40
20


Percentage Confusion Matrix: qda
92.67
5.02
1.67
0.64
True Labels
0
1
0
1
Predicted Labels
80
60
40
20


Percentage Confusion Matrix: random_forests
97.69
0.00
0.49
1.83
True Labels
0
1
0
1
Predicted Labels
80
60
40
20
0


Percentage Confusion Matrix: svm
97.69
0.00
2.31
0.00
True Labels
Predicted Labels
0
1
80
60
40
20
0


Percentage Confusion Matrix: decision_trees
96.45
1.23
2.08
0.23
True Labels
Predicted Labels
0
1
80
60
40
20

4. Discussion:

4.1 Challenges and Limitations:

Quadratic Discriminant Analysis (QDA) is a powerful statistical technique used for classification problems. Despite its advantages, QDA has several limitations that may affect its performance and suitability for some applications. The following are some of the limitations, along with recommendations for future research:

Normality assumption:
QDA assumes that the predictor variables are normally distributed within each class. If this assumption does not hold, model performance can be significantly compromised.

Sensitivity to extreme values:
Because QDA relies on the covariance of predictors, it can be sensitive to outliers. Outliers may have an excessive effect on the estimation of covariance matrices, leading to less reliable classification.

High dimensional data:
QDA can perform poorly when the number of features (dimensions) approaches or exceeds the number of observations, a situation known as the "curse of dimensionality". In such cases, the estimation of individual covariance matrices becomes less stable and accurate.

Computational complexity:
Estimating separate covariance matrices for each class can become computationally intensive, especially with a large number of predictor variables or classes.

Overfitting:
QDA can overfit the data, especially when the number of predictions is large relative to the sample size. This is due to the model's flexibility in fitting the data using quadratic decision bounds.

Class Imbalance:
In scenarios where there are many more observations in one category than another (common in bankruptcy prediction), the performance of QDA can be degraded as it does not inherently capture category imbalance.

Covariance structure:
QDA assumes that each class has its own covariance structure. This can be a limitation if the true covariance structures are not sufficiently distinct or if there is insufficient data to estimate them accurately.

4.2 Future Research:

Feature Selection and Dimension Reduction:
Applying feature selection techniques or dimensionality reduction methods such as PCA (Principal Component Analysis) before applying QDA could improve performance and computational efficiency.

Powerful QDA:
Developing more robust versions of QDA that are less sensitive to outliers and violations of the normality assumption may be a valuable area of research.

Arrangement techniques:
Investigating regularization techniques to avoid overfitting and more efficiently handle high-dimensional datasets could improve the utility of QDA.

Hybrid models:
Combining QDA with other machine learning approaches, such as ensemble or boosting methods, to improve predictive performance and stability.

Class Imbalance Management:
Investigate methods of adapting QDA to perform better on unbalanced datasets, such as incorporating class weights or using sampling techniques.

Alternative covariance estimate:
Investigation of alternative methods for estimating covariance matrices in QDA, which may provide more stable and accurate classification in some situations.

Cross-validation and model selection:
Applying rigorous cross-validation techniques to select the best model and tune hyperparameters can help mitigate overfitting and improve model generalization.

Application Specific Tuning:
Tailoring QDA to specific applications by incorporating domain knowledge into the modelling process can potentially improve its performance in these scenarios.

In summary, while QDA has some limitations, there are many opportunities for future research to improve its performance and application. Addressing its limitations through methodological improvements and adaptive techniques could extend its utility to various classification problems, including bankruptcy prediction.

5. Conclusion:

The extensive analysis of various machine learning models for predicting corporate bankruptcy underscores the potential effectiveness of Quadratic Discriminant Analysis (QDA). QDA's superior composite score, which adeptly balances the F1 and AUC ROC scores, highlights its proficiency. This balance is crucial in the intricate realm of financial distress forecasting. QDA's distinction lies in its capacity to model non-linear decision boundaries, facilitated by its allowance for distinct covariance structures across different classes. This feature enables QDA to adeptly capture complex patterns within the data, a significant advantage in the nuanced field of bankruptcy prediction.

Despite these strengths, QDA is not devoid of challenges. Its inherent assumptions, such as the normal distribution of data within each category, render it sensitive to outliers. This sensitivity can potentially lead to overfitting, especially in scenarios involving high-dimensional data spaces. Additionally, QDA's computational requirements can be demanding, necessitating careful consideration of its application.

Addressing these challenges necessitates future research efforts that could focus on several key areas. Robust estimation methods could mitigate the influence of outliers, enhancing QDA's resilience. Dimensionality reduction techniques might help in managing high-dimensional data, reducing the risk of overfitting and lessening computational demands. Regularization strategies could provide another avenue for preventing overfitting. Exploring hybrid models that combine QDA's strengths with those of other algorithms could result in models that are both robust and versatile. Furthermore, integrating domain-specific knowledge and adaptive methodologies tailored to particular contexts could significantly augment QDA's predictive accuracy.

The selection of an appropriate model demands a deep comprehension of each model's unique traits and the specific business context. This selection process involves a careful balance between accuracy and practical considerations, such as computational efficiency, model interpretability, and the economic impact of prediction errors. With judicious application and ongoing refinement, QDA, alongside other advanced machine learning models, holds substantial promise in offering insightful foresight into the early detection of business bankruptcy. This potential, when fully harnessed, could be pivotal in guiding critical business decisions and strategies in the ever-evolving financial sector

Reference List:

1. Qin, Y. (2018). A review of quadratic discriminant analysis for high-dimensional data. WIREs Computational Statistics, 10(4). (http://doi.org/10.1002/wics.1434). Accessed December 2023.

2. Schaffer, C. (1993). Selecting a classification method by cross-validation. Machine Learning, 13(1), 135–143. (http://doi.org/10.1007/bf00993106). Accessed December 2023.

3. Onyinyechi Jessica Egwom, Hassan, M., Jesse Jeremiah Tanimu, Hamada, M., & Oko Michael Ogar. (2022). An LDA–SVM Machine Learning Model for Breast Cancer Classification. BioMedInformatics, 2(3), 345–358. (http://doi.org/10.3390/biomedinformatics2030022). Accessed December 2023.

4. Barry de Ville. (2013). Decision trees. WIREs Computational Statistics, 5(6), 448–455. (http://doi.org/10.1002/wics.1278). Accessed December 2023.

5. Guo, G., Wang, H., Bell, D. A., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. Lecture Notes in Computer Science, 986–996. (http://doi.org/10.1007/978-3-540-39964-3_62). Accessed December 2023.

6. Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. Lecture Notes in Computer Science, 1015–1021. (http://doi.org/10.1007/11941439_114). Accessed December 2023.

7. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. (http://doi.org/10.1023/A:1010933404324). Accessed December 2023.
8. Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking & Finance, 18(3), 505-529. (http://doi.org/10.1016/0378-4266(94)90007-8). Accessed December 2023.

9. Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 18(1), 109-131. (http://doi.org/10.2307/2490395). Accessed December 2023.

10. Friedman, J. H. (1989). Regularized discriminant analysis. Journal of the American Statistical Association, 84(405), 165-175. (http://doi.org/10.1080/01621459.1989.10478752). Accessed December 2023.

11. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874. (http://doi.org/10.1016/j.patrec.2005.10.010). Accessed December 2023.

12. Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. European Journal of Operational Research, 180(1), 1-28. (http://doi.org/10.1016/j.ejor.2006.08.043). Accessed December 2023.

13. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. (http://doi.org/10.1007/978-0-387-84858-7). Accessed December 2023.

14. Powers, D. M. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies, 2(1), 37-63. (https://www.inderscience.com/info/inarticle.php?artid=21471). Accessed December 2023.

15. López de Prado, M. (2018). Advances in Financial Machine Learning. Wiley. ISBN 978-1-119-48208-6. (https://www.wiley.com/en-us/Advances+in+Financial+Machine+Learning-p-9781119482086). Accessed December 2023.