**Report: Credit Scoring with Boosted Decision Trees**

**Introduction**

This study compares the performance of four machine learning models for credit risk assessment using the German credit dataset. The models investigated are Decision Trees with Pruning, AdaBoost, Multilayer Perceptron (MLP), and Support Vector Machine (SVM). Evaluation of each model's performance is conducted through classification reports and statistical tests to compare their area under the ROC curve (AUC) estimates.

**Methods**

Data preprocessing involved handling missing values and partitioning the dataset into training and testing sets. Subsequently, each model was trained using either default or optimized hyperparameters determined through grid search and cross-validation. Evaluation metrics encompassed precision, recall, F1-score, and accuracy. Additionally, ROC curves were plotted for visual comparison.

**Results**

**Decision Trees with Pruning**

The pruned decision tree achieved an accuracy of 0.69. However, precision, recall, and F1-score for class 0 were comparatively lower than for class 1, indicating imbalanced performance.

**AdaBoost**

AdaBoost, with optimal parameters (learning rate = 1.0, n\_estimators = 300), achieved an accuracy of 0.78, surpassing the decision tree in precision, recall, and F1-score for both classes.

**Support Vector Machine (SVM)**

The MLP model, with optimized parameters (hidden\_layer\_sizes = (21,), learning\_rate\_init = 0.0001, max\_iter = 300), exhibited an accuracy of 0.69. However, it displayed poor performance for class 0, with zero precision and recall.

**Statistical Analysis**

The Wilcoxon-Mann-Whitney test was conducted to compare the AUC estimates of the models. The results revealed no significant difference in performance between any pair of models, supported by high p-values and low T-statistics.

**Discussion**

AdaBoost showcased superior overall performance, with the highest accuracy and balanced precision and recall for both classes. Decision Trees with Pruning and SVM also exhibited competitive performance, albeit with some class-specific metric imbalances. MLP, conversely, displayed inadequate performance, particularly for class 0, highlighting the necessity for further optimization or alternative modeling approaches. The statistical analysis reinforced these findings, indicating comparable performance among the models.

**Conclusion**

In conclusion, AdaBoost emerges as the preferred model for credit risk assessment in this study. However, addressing imbalanced performance in other models and exploring alternative techniques to enhance model accuracy and reliability warrant further investigation.