Project Report: Bankruptcy Prediction

Introduction: Bankruptcy prediction is a critical area of research in finance, where the goal is to develop models that can accurately predict the likelihood of a company going bankrupt. In this project, we use logistic regression and feature engineering techniques to predict bankruptcy.

Project Overview: We start by loading the data and cleaning it by getting rid of inconsistent nan values. Since our target variable is imbalanced, we use SMOTE resampling to balance it. Next, we create a logistic regression model and fit it with the resampled Bankruptcy column. We then perform feature engineering and select the 10 best features to improve our model's accuracy, which jumps from 0.72 to 0.86.

Implementation: We use Python and several libraries like Pandas, NumPy, and Scikit-learn to implement this project. We use the LogisticRegression module from Scikit-learn to build the logistic regression model and SMOTE from the imblearn library to balance the target variable. We also use feature selection techniques from Scikit-learn to select the best features.

Conclusion: Our project demonstrates that logistic regression combined with feature engineering is an effective way to predict bankruptcy. Our model achieved an accuracy of 0.86, which is a significant improvement from the original accuracy of 0.72. However, further improvements can be made by using more advanced techniques like ensemble learning or neural networks.

Future Work: In the future, we can explore other feature selection techniques and try other resampling methods to see if they can improve our model's performance.