<u>Report</u>

- This report delves into our methodology for predicting financial distress through the application of diverse balancing strategies.
- We explore how we handled the imbalance in the dataset and analyze the effects of these strategies on the performance of our predictive models.

<u>Satisfaction with Findings:</u>

1) Balancing Class Distribution:

<u>Undersampling:</u> While effectively mitigating the minority class imbalance, this approach allowed for less data volume and revealed issues with underfitting.

<u>Oversampling:</u> Oversampling generated promising results, particularly in precision. However, recall declined a little when compared to SMOTE.

<u>SMOTE:</u> As a result of its ability to synthesize synthetic samples, SMOTE was able to significantly improve model generalization, highlighting its effectiveness for recognizing patterns underneath.

2) Results of the Model:

- In comparing with the methodologies of Undersampling, the Oversampling and SMOTE methods demonstrated superior performance, as determined by their elevated F1 Scores.
- In comparing with SMOTE, oversampling demonstrated extraordinary precision, nevertheless with a slightly lower recall capacity.
- SMOTE accomplished remarkable results in terms of generalization and stability, thereby establishing itself as the go-to option for robust modeling.

These understandings deliver additional verification that these strategies are effective in enhancing model performance while dealing with challenges connected with class imbalance.

3) Possibilities for Development:

The recognition of potential possibilities for advancement consisted of the exploration of additional features through the methodology of feature engineering and the implementation of ensemble methods to achieve higher standards of predictive accuracy.

Significant Takeaways discovered from Each Stage:

- Preprocessing of the Data:

The crucial procedures in handling missing values and transforming the target variable lay down the foundation for future evaluations.

- Balancing Methods:

Recognizing the distinct impacts of undersampling, oversampling, and SMOTE on model performance demonstrated the importance of choosing the best balancing strategy based on dataset characteristics along with desired outcomes.

- Model Training and Analysis:

Knowledge into the advantages and drawbacks of models trained on balanced datasets provided useful guidance for model selection and optimization, taking consideration of variations in precision, recall, and F1 score.
The comprehensive analysis supplied practical knowledge into class imbalance handling and suggested various possibilities for model enhancement and improved performance.

Visualizations of Confusion Matrix: