# ML Lab 06

## Random Forest Model for Classification of an Imbalanced Dataset

After applying the SMOTE (Synthetic Minority Over-sampling Technique) method to handle class imbalance, the model's evaluation metrics have improved significantly:

* Precision increased from 0.4486 to 0.5191 - This indicates that the model now makes fewer false positive predictions, meaning it's better at correctly identifying the positive class.
* Recall decreased from 0.9289 to 0.8103 - While recall has dropped, this suggests the model is now more balanced and not overly biased toward predicting the majority class.
* F1 Score improved from 0.6051 to 0.6328 - This confirms an overall better balance between precision and recall.
* Accuracy increased from 0.8634 to 0.8941 - A noticeable improvement in overall model performance.

The initial model had very high recall but low precision, meaning it was predicting the most positives correctly but also had a high false positive rate. After handling class imbalance with SMOTE, the model achieved a better balance between precision and recall, leading to a more reliable classification. The increase in accuracy and F1 score suggests that oversampling helped the model generalize better, rather than just overfitting to the majority class.

## Random Forest Model for Regression

A graph with a line and a number of trees

AI-generated content may be incorrect.

* Rapid Decrease Initially - When the number of trees increases from 1 to around 25, there is a sharp drop in both training RMSE (blue line) and test RMSE (orange line). This indicates that the model is improving and reducing errors significantly.
* Plateauing Effect - After 50 trees, the test RMSE stabilizes around 0.5, and further increasing trees does not improve test performance. Training RMSE continues to decrease slightly but flattens out around 0.15-0.2.
* Overfitting Behavior - The training RMSE is much lower than the test RMSE, suggesting some degree of overfitting. The model fits the training data very well but does not generalize as effectively to the test set.

**What Happens if Trees Increase Further?**

* No significant reduction in test RMSE - Additional trees do not improve test accuracy.
* Higher computational cost - Training and inference become slower.
* Diminishing returns - More trees won't meaningfully impact performance beyond an optimal number (~150-175 trees in this case).