# Model Optimization and Hyperparameter Tuning

**Hyperparameter Tuning Report (PDF)**

**Overview:**

In this report, I take a deep dive into the fine-tuning of my machine learning models to ensure they perform at their best. By adjusting key parameters, I can optimize how well the model fits the data and ultimately improve accuracy, reliability, and efficiency.

**Key Sections:**

- **Tuning Methods and Strategy:**
  - I experimented with a variety of methods to identify the best hyperparameter settings, including Grid Search, where I systematically tried out combinations of parameters, and Random Search, which helped save time while covering a wide range of options.
- **What I Found:**
  - I present detailed results showing how the best combination of hyperparameters influenced model performance. The tables highlight the accuracy, precision, recall, and F1 score for the selected model configuration, showcasing its optimal performance across these key metrics.
- **Impact of Hyperparameters:**
  - For each hyperparameter, I explored its impact on the model's behaviour. For instance, tweaking the max_depth parameter influenced how complex the decision-making process was, while adjusting n_estimators in Random Forest affected both speed and accuracy.
  - I provided a clear understanding of what each setting does and how it balances between underfitting and overfitting the model.
- **Choosing the Best Model:**
  - Based on these results, I picked the most effective model configuration. This model was selected for its strong performance across all metrics while maintaining interpretability and computational efficiency.

**Model Interpretability Report**

**Why This Matters:**

Understanding how my models make decisions is just as important as ensuring they perform well. This report focuses on explaining the model's predictions in a way that makes sense, providing transparency and helping you trust the results.

**What's Included:**

- **Feature Importance:**
    - I start by looking at which features have the most impact on predictions. Through feature importance plots, you can see which factors (like age or academic performance) play a crucial role in the model's decisions. For example, the Admission Grade consistently emerged as one of the strongest predictors.
- **SHAP (SHapley Additive exPlanations):**
    - To dig deeper into model explanations, I used SHAP values, which break down the contribution of each feature for individual predictions.
        - **SHAP Summary Plot:** A visualization that shows the overall importance and direction (positive or negative) of each feature.
        - **SHAP Dependency Plots:** These plots offer insights into how specific features interact with predictions, allowing me to see, for instance, how an increase in Age might affect the model's output.
- **Partial Dependence Plots (PDP):**
    - I've also included PDPs, which help explain the relationship between important features and model predictions. For instance, if I look at Curricular Units, the PDPs can show how different levels of student enrollment influence the likelihood of certain outcomes.
- **Wrapping It Up:**

- At the end of the report, I summarize the key takeaways from the interpretability analyses and the final model evaluation results. The goal was to not only understand which features matter most but also to identify the most suitable model for our use case, balancing performance, interpretability, and computational efficiency.

- **Model Evaluation Summary:**
    - After evaluating multiple models, I found that **XGBoost** stood out, delivering the highest accuracy (0.8881), precision (0.8776), recall (0.7570), and F1 score (0.8129). Its confusion matrix showed a strong ability to correctly identify both true positives and true negatives.

- While the **Random Forest** also demonstrated good performance (accuracy of 0.8802 and precision of 0.8836), its recall was slightly lower, leading to fewer positive cases being captured.
- The **Decision Tree**, on the other hand, offered ease of interpretability but fell behind in terms of overall accuracy (0.8316) and F1 score (0.7256).

- **Final Model Selection:**
    - **XGBoost** was selected as the final model due to its superior performance metrics across the board. While it is less interpretable than a Decision Tree, it provides good insights into feature importance, making it a practical choice.
    - If interpretability had been the top priority, the **Decision Tree** could have been considered for its transparency. However, given the objective of achieving the best overall performance, **XGBoost** is the optimal choice.